Identification of developmental functions for *Arabidopsis thaliana* genes by a reverse genetics approach based on analysis of H3K27me3 distribution

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Julia Carina Engelhorn
aus Haan

Köln, 2011

Die vorliegende Arbeit wurde am Max-Planck-Institut für Züchtungsforschung in Köln in der Abteilung für Entwicklungsbiologie der Pflanzen (Direktor Prof. Dr. G. Coupland) angefertigt.

Prüfungsvorsitzender:	Prof. Dr. Martin Hülskamp
Berichterstatter:	Prof. Dr. George Coupland
	Prof. Dr. Wolfgang Werr
Tag der Disputation:	2. Februar 2011

Abstract

Polycomb Group (PcG) protein mediated gene repression is essential for normal development in both plants and animals, as demonstrated by severe developmental defects resulting from their loss-of-function. PcG proteins convey repression of target genes by tri-methylation of lysine 27 of histone 3 (H3K27me3). Many H3K27me3 decorated genes encode developmental regulators in *Arabidopsis thaliana* and developmental functions are particularly overrepresented in tissue specific sub sets of H3K27me3 targets. This study identified 105 genes specifically expressed in the shoot apex and floral organs by transcriptional clustering analysis, which are particularly enriched for shoot developmental functions according to Gene Ontology analysis. As half of the genes in this group were not characterised in detail, these were screened for a role in shoot development by analysing loss-of-function mutants and selected candidate gene overexpessor plants. Fourteen putative Development related PcG Targets in the Apex (DPAs) were identified. For five DPA putants developmental abnormalities were confirmedly associated with the respective loci. Among them were genes related to flowering time, leaf size and leaf shape regulation.

dpa4 loss-of-function plants display enhanced leaf serrations and enlarged petals, while leaf margins of *35S::DPA4* plants are smooth. *DPA4* encodes for a putative RAV (Related to ABI3/VP1) transcriptional repressor and is expressed in the lateral organ boundary region and in leaf sinuses. Total leaf area and cell numbers are not altered in *dpa4* plants, suggesting that DPA4 regulates leaf margin outgrowth by inhibiting growth towards leaf serrations. *DPA4* expression domains widely overlap with those of *CUP-SHAPED COTYLEDON 2*, known to regulate leaf margin shape. Genome-wide transcriptional profiling in *dpa4* apices revealed 77 differentially expressed genes. An overrepresentation of auxin-response elements in the promoters of these otherwise poorly characterised genes indicates a role for DPA4 in auxin-mediated signalling. This is further supported by an auxin-influx carrier mutant-like phenotype observed for *35S::DPA4* plants displaying left-hand twisted rosette leaves. Taken together, the data confirm that DPA4, which was identified as a candidate by this reverse genetics screen, is a newly identified player in the signalling network controlling leaf serrations in *Arabidopsis thaliana*.

Zusammenfassung

Die von Proteinen der Polycomb Gruppe (PcG) vermittelte Repression von Genen ist sowohl für die Entwicklung von Pflanzen als auch von Tieren essentiell. Ein Verlust dieser Repression hat schwerwiegende Entwicklungsdefekte zur Folge. PcG Proteine vermitteln die Repression von Zielgenen über eine Trimethylierung von Lysin 27 an Histon 3 (H3K27me3). Viele mit H2K27me3 markierte Gene kodieren Entwicklungsregulatoren in *Arabidopsis thaliana* und Entwicklungsfunktionen sind besonders in kleinen, gewebespezifisch expremierten Untergruppen überrepräsentiert. In der vorliegenden Arbeit wurden durch transcriptionelle Cluster Analyse 105 Gene ausgewählt, die spezifisch im Apex und in floralen Organen expremiert und beruhend auf einer Gene Ontology Analyse besonders für Entwicklungsfunktionen angereichert sind. Da die Hälfte der Gene aus dieser Gruppe nicht detailliert charakterisiert waren, wurden sie mit Hilfe von Funktionsverlust-Mutanten und Überexpressoren der Kandidaten bezüglich einer Rolle in der Sproßentwicklung untersucht. Es wurden 14 mutmaßliche entwicklungsbezogene PcG Zielgene (Englisch: Development related PcG Targets in the Apex (DPAs)) identifiziert. Für fünf davon konnte ein Assoziation der Entwicklungsabnormalitäten mit dem jeweiligen Locus bestätigt werde. Darunter waren Gene, die im Zusammenhang mit Blühzeitpunktsteuerung und Blattgrößesowie Blattformregulierung stehen.

Funktionsverlust-Mutanten von *DPA4* zeigen verstärkte Einkerbungen der Blätter und vergrößerte Blütenblätter, während die Blattränder in *35S::DPA4* Pflanzen glatt sind. *DPA4* kodiert einen mutmaßlichen RAV Transkriptionsrepressor und wird in der Grenzregion zu lateralen Organen und im Sinus von Blättern expremiert. Blattfläche und Zellzahl sind in *dpa4* insgesamt nicht verändert, was darauf hindeutet, dass DPA4 den Auswuchs des Blattrandes durch eine Hemmung des Wachstums zu den Blattzähnen hin reguliert. Die Expressionsdomänen von *DPA4* überlappen zu großen Teilen mit der von *CUP-SHAPED COTYLEDON 2*, einem bekannten Regulator für Blattrandform. Eine genomweite Transcriptionsanalyse in *dpa4* Apices zeigte 77 differentiell expremierte Gene. Eine Überrepräsentierung von "Auxin-Antwort Elementen" (auxin-response elements) in den Promotoren dieser ansonsten wenig charakterisierten Gene weißt auf eine Rolle von DPA4 im Auxin-Signalweg hin. Dies wird ebenfalls durch Ähnlichkeiten zwischen *35S::DPA4* Pflanzen und Mutanten des Auxin-Influx Systems unterstützt, die beide linkshändig gedrehte Rosettenblätter aufweisen. Zusammengenommen bestätigen die Daten DPA4, das in diesem reversen genetischen Screen als Kandidat identifiziert wurde, als einen neuen Beteiligten im Signalweg der Kontrolle von Blatteinkerbungen in *Arabidopsis thaliana*.

Contents

1 Introduction

1.1 Role of histone modifications in the development of multicellular organisms

Development of multicellular organisms requires the establishment of stage and tissue specific gene expression patterns to enable differentiation and specification of distinct cell types. An important mechanism to establish cell specific expression is the stable repression of genes from early embryogenesis on that is only released in a certain tissue, stage or as response to an external stimulus. Stable repression of genes can be achieved by compaction of chromatin, which prevents RNA polymerases from transcribing a locus.

Eukaryotic DNA is usually organised in a higher order structure named chromatin. This organisation is mediated by histone proteins. The histone core protein complex consists of four hetero-dimers between the histone proteins histone 3 (H3) and histone 4 (H4) and histone 2A and 2B (H2A and H2B). A 1.65 superhelical turn of 146 base pairs (bp) of DNA together with this histone complex form the core nucleosome. Nucleosomes occur on average every 200 bp and are further compacted by a fifth histone, histone 1 (H1), by stabilisation of higher order structures and fixation of the core histone complex to the DNA (McGhee and Felsenfeld, 1980; Widom, 1989; Luger et al., 1997).

The compaction of the DNA facilitates packing of the large DNA molecules into the nucleus, but also allows a mechanism of transcriptional regulation independent of the DNA sequence: initiation of transcription requires binding of RNA polymerases and transcription factors to the DNA and this binding can be hindered by nucleosomes. Nucleosomes can either mask the binding site of transcription factors or simply prevent the assembly of large transcription initiation complexes. Furthermore, it is necessary to remove at least the H2A/H2B dimers from the nucleosomes during transcript elongation by RNA Polymerase II (Li et al., 1997; Thiriet and Hayes, 2006). Therefore, transcription levels can be regulated by compaction and loosening of chromatin and by nucleosome positioning.

According to the “histone code” hypothesis, positioning of histones and chromatin

state can be controlled by information contained in covalent modifications of histone tails (Strahl and Allis, 2000; Jenuwein and Allis, 2001). The amino-terminal tails of histones are not condensed into the core nucleosome particle and are accessible for modifying proteins (Luger et al., 1997). Histone marks can be either associated with repression or activation of transcription: acetylation and phosphorylation of histones are rather associated with activation, while methylation and ubiquitination of histones can correlate to activation or repression of the associated genes, depending on the amino acid residue that is modified (Berger, 2007). Histone modifications are stable during mitosis and therefore provide an epigenetic mechanism for stable gene regulation over several rounds of cell cycling.

Chromatin can be subdivided in a gene-rich part, named euchromatin, which is less condensed in interphase nuclei and the very condensed, gene-poor heterochromatin (Grewal and Elgin, 2007). These two varieties of chromatin are associated with different predominant histone marks. In accordance to this, genes belonging to heterochromatic regions that carry heterochromatin associated marks are stably silenced and unlikely to be reactivated, while euchromatic genes can be repressed for a certain period during development and afterwards activated. An example for different repressive euchromatic and heterochromatic histone modification is the methylation of lysine residues in H3. In animals and plants, H3K27 tri-methylation (H3K27me3) mediates reversible repression of euchromatic genes, while H3K9 di-methylation (H3K9me2) is associated with stably silenced heterochromatic regions (Jenuwein and Allis, 2001; Ringrose and Paro, 2004).

1.1.1 Mechanism of H3K27me3 mediated repression

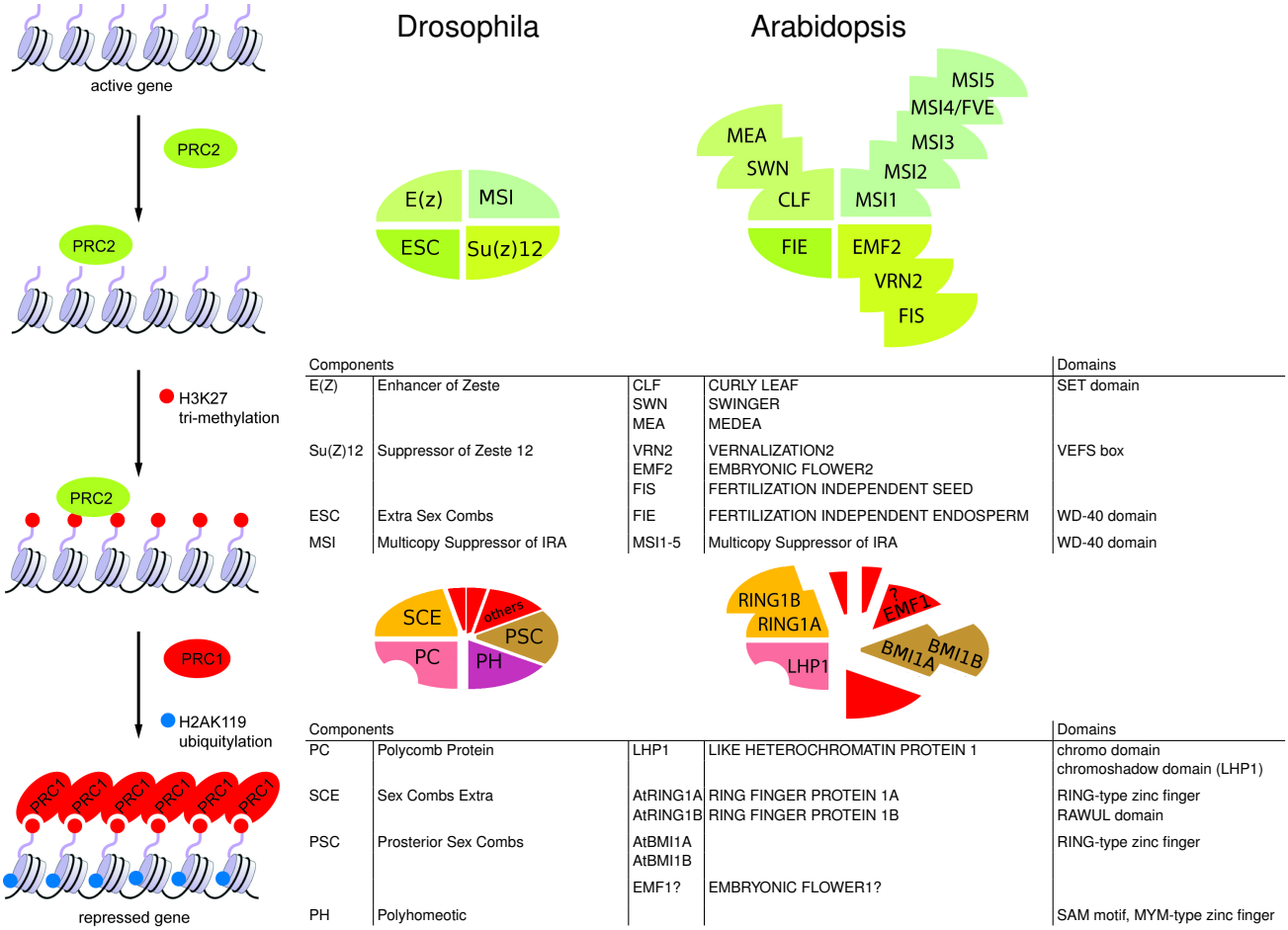
H3K27me3 is associated with the Polycomb Group (PcG) proteins, which are at least partially conserved in all multicellular organisms (Schwartz and Pirrotta, 2008). In insects and mammals, PcG proteins assemble in several Polycomb Repressive Complexes (PRC1-PRC4) which are involved in the recognition and stable repression of target genes via methylation of H3K27. PRC1 and PRC2 are the major complexes, PRC3 and PRC4 are variants of PRC2. PRC2 methylates H3 at genes that are to be repressed and the modification is subsequently recognised by PRC1, which leads to maintenance of the repression (Kuzmichev et al., 2005; Schwartz and Pirrotta, 2008) (Fig. 1.1). The molecular

mechanism leading to repression is not yet completely known, but it has been shown that PRC1 stays associated with the silenced locus and catalyses mono-ubiquitylation of lysin 119 in H2A, which then leads to a compaction of chromatin and prevention of transcription initiation (Morey and Helin, 2010). Once the PRC2 is recruited to a locus, the methylation mark is usually broadly spread over the region, often covering several genes. The spreading can be prevented by other chromatin marks such as the heterochromatic mark H3K9me2 (Turck et al., 2007; Farrona et al., 2008). In *Drosophila melanogaster* (Drosophila), cis-regulatory regions have been discovered that target PRCs to the genes to be repressed. These so-called Polycomb group response elements (PREs) are varying in their sequence and no consensus sequence has been identified up to now. A common mechanism among the PREs seems to be the binding capability to Polyhomeotic (PH), a PRC1 member from Drosophila, suggesting that PRC1 is involved in the recognition of target genes. However, no PREs have been identified in plants and mammals so far but several recent reports suggested non-coding RNAs (ncRNAs) to interact with PRC2 and recruit the complex to target loci in mammals (Simon and Kingston, 2009).

1.1.2 Composition of PRCs in Drosophila

The eponymous protein for the PcG is the chromodomain containing protein Polycomb (PC), which is a part of PRC1 and was first discovered in Drosophila. PC binds to H3K27me3 and is thereby involved in the recognition of this histone mark. Other components of PRC1 in Drosophila are PH, Posterior Sex Combs (PSC) and the Ring finger protein Sex Combs Extra (SCE). PSC is involved in complex formation and the inhibition of transcription, while RING proteins catalyse mono-ubiquitylation of lysin 119 in histone 2A (H2A) (Fig. 1.1).

Drosophila PRC2 consists of four Proteins: the SET (SUVAR3-9/E(Z)/Trithorax) domain protein Enhancer of Zeste (E(Z)) is the catalytic core and confers the methyl transferase activity in which it is aided by Suppressor of Zeste 12 (Su(Z)12), Extra Sex Combs (ESC) and Multicopy Suppressor of IRA (MSI) (Morey and Helin, 2010)(Fig. 1.1).



Components				Domains
E(Z)	Enhancer of Zeste	CLF SWN MEA	CURLY LEAF SWINGER MEDEA	SET domain
Su(Z)12	Suppressor of Zeste 12	VRN2 EMF2 FIS	VERNALIZATION2 EMBRYONIC FLOWER2 FERTILIZATION INDEPENDENT SEED	VEFS box
ESC	Extra Sex Combs	FIE	FERTILIZATION INDEPENDENT ENDOSPERM	WD-40 domain
MSI	Multicopy Suppressor of IRA	MSI1-5	Multicopy Suppressor of IRA	WD-40 domain

Components				Domains
PC	Polycomb Protein	LHP1	LIKE HETEROCHROMATIN PROTEIN 1	chromo domain chromoshadow domain (LHP1)
SCE	Sex Combs Extra	AtRING1A AtRING1B	RING FINGER PROTEIN 1A RING FINGER PROTEIN 1B	RING-type zinc finger RAWUL domain
PSC	Prosterior Sex Combs	AtBMI1A AtBMI1B		RING-type zinc finger
		EMF1?	EMBRYONIC FLOWER1?	
PH	Polyhomeotic			SAM motif, MYM-type zinc finger

Figure 1.1: **Mechanism of PcG-mediated gene repression and composition of PRCs in plants and animals.** PRC2 catalyses tri-methylation of target loci at H3K27 via SET domain proteins (left part of the figure). Components of PRC2 are conserved between plants and animals, but are encoded by only four genes in Drosophila while all *Arabidopsis thaliana* components except FIE are encoded by small gene families (right part of the figure). H3K27me3 is recognised by PRC1 in Drosophila and catalyses mono-ubiquitylation of lysine 119 in H2A, which then contributes to stable repression of the target locus. In Arabidopsis LHP1, EMF1, AtRING1A and B and AtBMI1A and B are believed to confer PRC1 function. Recently H2AK119 ubiquitylation activity was reported for AtBMI1A and B (Bratzel et al., 2010). Figure modified after Adrian et al. (2009).

1.2 Polycomb Group protein mediated gene repression in plants

In plants, PRC2 is completely conserved and proteins homologous to all PRC2 members in Drosophila were identified. PRC1 function also seems to be conserved, but not all members of the plant PRC1 that are known so far are homologs to Drosophila proteins (Pien and Grossniklaus, 2007).

1.2.1 PRC2 components in Arabidopsis

In *Arabidopsis thaliana* (Arabidopsis), the PRC2 complex consists of several partially redundant components that assemble in different complexes (Fig. 1.1). Several target genes are repressed by only one complex but others are common targets of different complexes (Makarevich et al., 2006). Su(Z)12 homologs in Arabidopsis are FERTILIZATION INDEPENDENT SEED 2 (FIS2), VERNALIZATION 2 (VRN2) and EMBRYONIC FLOWER 2 (EMF2). Depending on the contained Su(Z)12 homolog, the complex can for example rather function in establishing vernalization (cold treatment to induce flowering) response (VRN2 complex) or in regulation of photoperiodic flowering (EMF2 complex) (Farrona et al., 2008).

The catalytic core of PRC2 is made up of one of three partially redundant SET domain proteins CURLY LEAF (CLF), SWINGER (SWN) and MEDEA (MEA), which are E(Z) homologs. Expression of *MEA* is restricted to the gametophyte and the endosperm and is therefore part of a gametophyte specific PRC2, while CLF and SWN act in the sporophyte. In accordance, *mea* loss-of-function plants show aborted development of gametophyte and endosperm, *clf* mutant plants display developmental defects like homeotic transformations of floral organs, early flowering and curled leaves, while *swn* plants develop completely normal (Grossniklaus et al., 1998; Chanvivattana et al., 2004; Schubert et al., 2005). In the double mutant of the partially redundant components *CLF* and *SWN*, cell differentiation is strongly disturbed, resulting in a callus-like structure rather than a differentiated plant (Schubert et al., 2005).

Five MSI homologs (MSI1-5) exist in Arabidopsis, but only MSI1 is confirmedly involved in PRC2 complexes. MSI1 is involved in seed development as part of the MEA containing PRC2; seeds carrying a homozygous mutation of *MSI1* are aborted (Köhler et al., 2003). There is only one ESC homolog, the WD-40 domain protein FERTILIZATION INDEPENDENT ENDOSPERM (FIE) (Adrian et al., 2009; Farrona et al., 2008). Therefore, this component will be common to all PRC2 complexes and *fie* mutant female gametophytes, that initiate endosperm development without fertilisation display complete lethality (Ohad et al., 1999).

1.2.2 Putative PRC1 members in Arabidopsis

Six proteins have been suggested in the last years to be part of a complex with PRC1 function in plants:

TERMINAL FLOWER 2/LIKE HETEROCHROMATIN PROTEIN 1 (TFL2/ LHP1) is the only Arabidopsis protein with overall sequence similarity to HETEROCHROMATIN PROTEIN 1 from metazoans and was therefore believed to be involved in heterochromatic repression of genes. However, determination of LHP1 target genes revealed a co-localisation of the protein with the H3K27me3 mark and hence a localisation to euchromatin (Turck et al., 2007; Zhang et al., 2007b). Since LHP1 is directly involved in the repression of some of the target genes of PRC2 components (Kotake et al., 2003), it could be a part of a non-conserved PRC1-like complex. This is further confirmed by the finding, that in the case of the H3K27me3 target *FLOWERING LOCUS T* (*FT*), H3K27me3 is not able to mediate repression alone but needs the presence of LHP1 (Adrian et al., 2010).

EMBRYONIC FLOWER1 (EMF1) participates in EMF2 mediated repression of the flower homeotic gene *AGAMOUS* (*AG*) and directly interacts with MSI1 *in vitro*. This suggests a PRC1 or PRC2 like function of EMF1. The fact that EMF1 was shown to interfere with *in vitro* transcription suggests a similarity to PSC and led to the conclusion that EMF1 is more likely to be part of PRC1 (Calonje et al., 2008).

Five RING-finger homologs have been identified in Arabidopsis, AtRING1A/B and AtBMI1A-C. RING1A and RING1B are the closest homologs of SCE in Arabidopsis and interact with LHP1, EMF1 and CLF *in vitro*. Therefore, they are suggested to be part of a PRC1-like complex in Arabidopsis (Sanchez-Pulido et al., 2008; Xu and Shen, 2008). Recently, AtBMI1A and AtBMI1B were shown to bind to LHP1 and EMF1 *in vitro* and to be able to monoubiqutinate H2A, thus exhibiting expected PRC1 functions (Bratzel et al., 2010). For AtBMI1C no such function could be shown so far.

Several existing double mutants of putative PRC1 members and PRC2 components support the existence of a PRC1-like complex in Arabidopsis. *AtRING1A*, *AtRING1B*, *AtBMI1A* and *AtBMI1B* single loss-of-function mutants display normal phenotypes, while the double mutants *Atring1a/Atring1b* and *Atbmi1a/Atbmi1b* display severe developmental defects. Defects observed in *Atring1a/Atring1b* plants are ectopic tissue formation, alterations in leaf shape, elevated numbers of floral organs and fasciated stems, while

Atbmi1a/Atbmi1b mutant seedlings turn into a callus like structure resembling the phenotype of *clf/swn* plants (Xu and Shen, 2008; Bratzel et al., 2010).

Loss of LHP1 function mutants flower early under both long day (LD) and short day (SD) conditions and form a terminal flower (Kotake et al., 2003). The differences in severeness of phenotypic changes suggest that some of the PRC1-like components act redundantly, which was shown for LHP1 and AtRING1A and AtRING1B. Triple *Atring1a/Atring1b/lhp1* or *Atring1a/Atring1b/clf* mutant plants show enhancement of the mutant phenotype, they are dwarfish and inflorescences arrest completely (Xu and Shen, 2008). The triple mutation including *lhp1* is stronger affected, in this plant even cotyledons do not expand normally, suggesting a stronger overlap of AtRING1A and AtRING1B function with LHP1 than with CLF. Since the triple *Atring1a/Atring1b/lhp1* mutant does not form a callus like structure as *Atbmi1a/Atbmi1b*, it seems that either AtBMI1A and AtBMI1B can compensate for AtRING1A, AtRING1B and LHP1 function. Optionally, AtBMI1A and AtBMI1B might target different genes which are not affected by the other components. The second possibility is likely to be true since repression of AG requires both CLF and LHP1 but not AtRING1A and AtRING1B (Kotake et al., 2003; Xu and Shen, 2008). This indicates that putative PRC1 components of Arabidopsis might also form complexes of different composition that target specific sub sets of genes.

1.3 Shoot development in Arabidopsis

Unlike animals, plants develop and grow throughout their whole life time. This is possible because of a retention of undifferentiated cells during plant development in structures termed meristems, in contrast to the fast differentiation of animal cells during embryonic development. Pluripotent stem cells contained in the meristem provide cells for the formation of different organs in the course of the plant's life , while in animals all organs are initiated during embryogenesis and multipotent stem cells are only retained in certain tissues like the bone marrow (Dodsworth, 2009; Malgieri et al., 2010).

All aerial organs of a plant emerge from the shoot apical meristem (SAM) located at tips of the shoot (shoot apex). This pool of undifferentiated cells is generated during embryogenesis and maintained throughout the life of the plant, enabling plants to adapt organ development to environmental cues such as light supply and thereby compensates

for their lack of mobility (Barton, 2010). Thus, developmental processes have to be tightly coordinated with environmental circumstances to ensure survival and reproductive success. In particular, annual plants like Arabidopsis are dependent on a precise timing of developmental processes such as flowering, since the plant will die after flowering and needs to reproduce in one season. Complex regulatory genetic networks process environmental cues and internal signals to ensure that the best timing for developmental phase transitions is achieved (Boss et al., 2004).

1.3.1 Developmental phases

Development of the Arabidopsis shoot can be divided into four main phases: juvenile vegetative phase, adult vegetative phase, reproductive phase and senescence (Lawson and Poethig, 1995; Bleecker and Patterson, 1997).

Juvenile phase

During the juvenile phase, a plant is unable to respond to environmental factors triggering the transition from the vegetative to the reproductive growth, meaning that a juvenile plant will not flower, even under inductive conditions. In Arabidopsis, the juvenile phase is short and anatomical attributes indicating juvenility are the formation of round, small leaves and a lack of abaxial trichomes (Telfer et al., 1997).

Adult vegetative phase

Adult plants growth vegetativly for a time controlled by environmental influences such as temperature and day length. Two different strategies are employed by Arabidopsis: winter annual accessions germinate in autumn, stay vegetative throughout the winter and flower in spring. Summer annuals or "rapid cyclers" germinate in spring and flower after a short vegetative phase in summer. The two widely used accession Columbia (Col-0) and Landsberg *erecta* (L*er*) belong to the later category (Grennan, 2006). Summer annuals have no vernalization requirement and flower fast (about four weeks from germination and with about 10 to 12 leaves) under inductive long day conditions. During the vegetative phase, leaves are produced to ensure supply of the plant with photosynthetically produced nutrients. To fulfil this purpose, the arrangement of leaves (and all following

lateral organs) is regulated in a way to obtain maximal incidence of light on each leaf: each leaf is initiated at the maximum distance from all previous leaves (137.5° average distance). This phenomenon is known as phyllotaxis (Kuhlemeier, 2007).

Reproductive phase

After the transition from vegetative to reproductive phase the plant no longer produces rosette leaves but starts to produce an elongating stem (bolting) with a certain number of side shoots produced in the axil of one or more cauline leaves. The SAM and the side shoot apical meristems change into an inflorescence meristem that produces flowers which give rise to fruits instead of leaves. The inflorescence meristems remain indeterminate and continuously produce flowers until the plant dies. Usually the main shoot grows out first and is the highest shoot that is never reached by side shoots (apical dominance) (Leyser, 2005).

Senescence

Once seed setting is started, a high amount of energy is needed to develop mature seeds and supply them with storage nutrients. Therefore, starting from the vegetative part of the plant, tissues undergo senescence and their nutrients are used for seed filling. Phenotypically, senescence is associated with yellowing of the leaves indicating chloroplast break down. Finally, all parts of the plant undergo senescence and the plant dies. The whole process needs to be highly coordinated to reach a maximum number of fully developed seeds (Bleecker and Patterson, 1997).

1.3.2 The role of the shoot apical meristem in development

Organisation of the SAM

Three zones can be distinguished in the SAM: a central zone (CZ), where pluripotent stem cells divide slowly, a peripheral zone (PZ), where the daughter cells of the CZ cells divide more rapidly and differentiate to produce new organs and a rib zone (RZ) underneath the CZ giving rise to the pith of the stem. The CZ can be divided into three outer layers L1-3 containing the stem cells and an organising center (OC) that ensures meristem maintenance (Fig. 1.2 A). To ensure persistence of the meristem, differentiation of

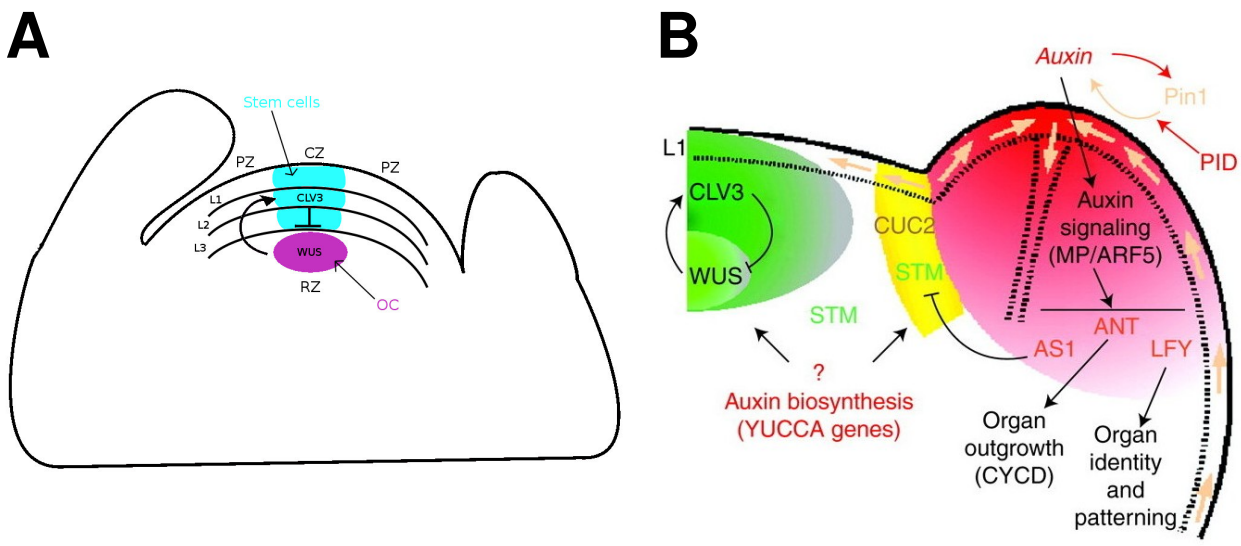


Figure 1.2: **Organisation of the SAM in Arabidopisis** **A**: Division of the SAM in central zone (CZ), peripheral zone (PZ) and rib zone (RZ). CZ is divided in 3 layers L1-3 and the organising center (OC). A feedback loop between *WUS* and *CLV3* ensures meristem maintenance. Figure modified from Miwa et al. (2009) **B**: Boundary specification and organ initiation in the shoot apex: organ initiation is antagonised by meristem identity genes *CUC2* and *STM* and promoted by *AS2*. Location of organ primordia is characterised by an auxin maximum that is established by PIN1 and PID. Apricot arrows indicate putative auxin transport directions. Elevated concentrations of auxin lead to activation of ARFs and finally direct organ initiation and outgrowth. Organ identity is then regulated by factors such as *LEAFY* (*LFY*). Figure displays example of an inflorescence meristem and was modified after Vernoux et al. (2010).

cells and maintenance of an undifferentiated cell pool have to be tightly regulated. This regulation is mainly realised by a feedback loop between the *CLAVATA* (*CLV1-3*) genes and *WUSCHEL* (*WUS*). The homeobox gene *WUS* is expressed in the OC and specifies stem cell identity, whereas *CLV* genes are expressed in L1-3 of the CZ and promote organ initiation (Laux et al., 1996; Laufs et al., 1998; Schoof et al., 2000; Miwa et al., 2009).

Organ initiation

Tightly regulated organ initiation in Arabidopsis is a result of the balanced interaction of so-called meristem identity genes and repressive factors that switch off these genes to allow organ initiation. Thus a boundary between the SAM and the emerging organs is established (Fig. 1.2 B). Meristem identity genes involved in this process are the class I KNOTTED1-like homeobox (*KNOX*) genes *SHOOTMERISTEMLESS* (*STM*), *BREVIPEDICELLUS* (*BP*) and KNOTTED-LIKE FROM ARABIDOPSIS THALIANA 2 (*KNAT2*) and

the partially redundant plant specific NAC (NAM (NO APICAL MERISTEM)-ATAF1,2-CUC2 (CUP-SHAPED COTYLEDON 2)) domain transcription factor encoding genes *CUC 1-3* (Aida et al., 1997; Takada et al., 2001; Vroemen et al., 2003; Irish, 2008). The *M*YB transcription factor *A*SYMMETRIC LEAVES 1 (*AS1*) and the Lateral Organ Boundary Domain (LBD) protein *ASYMMETRIC LEAVES 2* (*AS2*) are required for downregulation of *KNOX* genes in leaf primordia (Ori et al., 2000; Byrne et al., 2000, 2002).

Organ primordium formation and outgrowth is promoted by the plant hormone auxin (indole-3-acetic acid, IAA). In the L1 layer of the SAM, auxin maxima are created by a positive feedback loop between the auxin efflux carrier PIN-FORMED 1 (PIN1) and the auxin influx carriers *AUXIN1* (*AUX1*) and *LIKE AUX1 1-3* (*LAX1-3*) at the maximum distance to other primordia to ensure leaf initiation according to phyllotaxis (Reinhardt et al., 2003; Barkoulas et al., 2007; Bainbridge et al., 2008). Localisation of PIN1 to specific membranes is regulated via phosphorylation of the protein by the PINOID (PID) serine-threonine proteine kinase (Michniewicz et al., 2007). Further factors predicted to regulating auxin abundance are auxin biosynthesis related flavin monooxygenases of the YUCCA (YUC) family (Vernoux et al., 2010).

Auxin concentrations are perceived by AUXIN RESPONSE FACTORS (ARFs) and AUX/-IAA (Auxin/Indole-3-acetic acid inducible) transcription factors (Leyser, 2006; Vernoux et al., 2010) that regulate transcription of downstream genes leading to organ formation and outgrowth. This regulation can be positive or negative, depending on domains contained in the transcription factors (Ulmasov et al., 1999). Examples for downstream genes are the lateral organ size controlling gene *ARGOS* (auxin-regulated gene involved in organ size) (Hu et al., 2003) and the downstream regulated gene *AINTEGUMENTA* (*ANT*) (Mizukami and Fischer, 2000). The AP2/ERF (APETALA2/Ethylene-Responsive element binding Factor) containing transcription factor ANT regulates organ size by positively influencing the number of cell divisions in developing organs. *ANT* expression is both directly (by ARF2) and indirectly (through ARGOS) regulated by ARF transcription factors (Schruff et al., 2006).

Genetic networks controlling the vegetative to reproductive phase transition

During the vegetative phase, the SAM gives rise to leaves and lateral meristems. This production is stopped during the transition to the reproductive phase and flowers are pro-

duced. One key-player specifying floral meristem identity is the floral integrator *LEAFY* (*LFY*): its expression rises upon floral induction and leads to the production of flowers instead of leaves (Weigel et al., 1992). This is achieved by upregulation of floral organ identity genes like *AG* and *APETALA3* (*AP3*). *AP3* is also upregulated by another floral meristem identity gene, *UNUSUAL FLORAL ORGANS* (*UFO*) (Parcy et al., 1998).

The decision for the reproductive phase change of the SAM is triggered by different signals. In Arabidopsis, flowering time is regulated by four main pathways: the photoperiod pathway, the vernalization pathway, the autonomous pathway and the Gibberelic Acid (GA) pathway (Boss et al., 2004). Activation of these pathways leads to activation of floral integrators like *FT* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), depending on the respective pathway, which then induce reprogramming of the shoot apical meristem by activating *e.g. LFY* expression. The floral integrators are either directly synthesised in the apex or the gene product moves to the apex.

To activate the photoperiodic pathway, light is perceived in the leaves by the photoreceptors of the phytochrome (PHYTOCHROME A and B (PhyA and PhyB)) and cryptochrome family (CRYPTOCHROME 1 and 2 (CRY1 and CRY2)). In response to light activation, PhyA, CRY1 and CRY2 stabilise CONSTANS (CO) protein, the main transcriptional activator of *FT* (Valverde et al., 2004). In the absence of light, CO protein is degraded by SUPRESSOR OF PHYA-105-1 (SPA1) and the E3 ubiquitin ligase CONSTITUTIVE PHOTOMORPHOGENESIS 1 (COP1) (Laubinger et al., 2006). PhyB destabilises CO upon red light induction in the morning (Valverde et al., 2004). A complex of the proteins GIGANTEA (GI) and FLAVIN-BINDING, KELCH REPEAT, F-BOX PROTEIN 1 (FKF1) regulate transcription of CO in a circadian clock dependent manner so that expression occurs only in the evening, the time of day were light is present in LD but not in SD conditions (Sawa et al., 2007). Therefore, only in LD conditions, when *CO* transcription and the stabilisation of the protein coincide, CO protein reaches a sufficient level to activate *FT* transcription. FT protein then travels from the leaves to the SAM and induces flowering (Turck et al., 2008; Fornara et al., 2010).

To prevent precocious flowering under inductive conditions, *FT* transcription is repressed by two AP2/ERF domain containing RAV (Related to ABI3/VP1) transcription factors TEMPRANILLO 1 (TEM1) and RAV2/TEM2 (Castillejo and Pelaz, 2008). Also another AP2/ERF transcription factors, *SCHLAFMÜTZE* (*SMZ*), could be shown to directly

downregulate *FT* expression by binding to the *FT* locus (Mathieu et al., 2009). *SMZ* is, among other AP2/ERF transcription factors regulated by *miR172*.

Signals of the vernalization pathway, which represses flowering until the plant experiences a prolonged cold period, are integrated via the floral repressor *FLOWERING LOCUS C* (*FLC*). Prior to vernalization, *FLC* expression is upregulated by *FRIGIDA* (*FRI*) and thereby represses *FT* and *SOC1* expression. Loss of FRI function causes natural variation in Arabidopsis and is responsible for the "rapid cycling" behaviour of Col-0 and L*er* (Grennan, 2006). Upon cold temperature, *FLC* expression is stably reduced by several factors including antisense transcripts and PcG-protein mediated repression through the VRN2 complex (Swiezewski et al., 2009).

The autonomous pathway is not regulated by external cues but promotes flowering late in development, when no inductive conditions have occurred. This signal is integrated through a downregulation of *FLC*. This downregulation is required to enable the plant to respond to internal floral promoting signals of the GA pathway (Simpson, 2004). GA induces flowering in SD conditions, through activation of *SOC1* and *LFY* (Blazquez et al., 1998).

Two additional pathways have been recently identified: the ambient temperature pathway and the aging pathway. In higher temperatures (about23°C), flowering in Arabidopsis is accelerated through the release of *FT* repression by the MADS box transcription factor SHORT VEGETATIVE PHASE (SVP) (Li et al., 2008a). The aging pathway activates flowering through SQUAMOSA PROMOTER BINDING LIKE (SPL) transcription factors, which are in turn downregulated by the microRNA *miR156*. The microRNA concentration decreases during development (Wang et al., 2009).

Recent studies indicate that the two processes of floral transition initiation and floral organ specification are at least partially regulated by common factors. One example is *AP2*, which was first described to act as a homeotic floral organ identity gene specifying sepal and petal identity by regulating *AG* expression (Bomblies et al., 1999). More recently, AP2 was shown to be involved in timing regulation of floral transition as part of a complex network of feedback loops in which AP2 positively regulates the expression of *miR156*, which represses the *AP2* regulator *miR172* (Yant et al., 2010).

1.4 Role of H3K27me3 in Arabidopsis development

The pleiotropic developmental defects observed in mutants of PRC members already indicate that H3K27me3 plays a crucial role in development. Genome wide determination of H3K27me3 target genes (Turck et al., 2007; Zhang et al., 2007a) revealed that about 15 % of the annotated genes are decorated with this mark. H3K27me3 mediated repression was found in all steps of plant development from embryogenesis to floral organ identity determination, since many important developmental regulators such as *STM*, *CUC2*, *FT*, *LFY*, *AG* were among the target genes. This and the fact, that the H3K27me3 targets are expressed at low levels and in a rather tissue specific pattern (Zhang et al., 2007a), suggests that PcG-mediated repression participates in establishing and maintaining a tissue specific expression pattern of important developmental players ensuring normal plant development.

Nevertheless, it has to be noted that the tissue specific expression patterns are not caused by PcG repression alone, since there are examples where tissue specificity is still stable although the expression level is not. One example is *FT*, where *FT* expression is elevated in *lhp1* background but is still restricted to its typical vascular expression pattern (Adrian et al., 2010). This finding is supported by recent observation of discrete expression patterns for several H3K27me3 targets in *clf/swn* tissue in our group (personal communication, Dr. Sara Farrona, MPIPZ, Cologne). Therefore, the nature of the causal connection between tissue specificity and H3K27me3 labelling of genes remains unclear and is subject to current research.

The importance of H3K27me3 target genes for development is further supported by an overrepresentation of developmental functional genes in the target gene list compared to the rest of the genome. This overrepresentation is even stronger for sub sets of H3K27me3 targets expressed in the same tissue. Especially one sub set of targets that is expressed in floral organs showed high enrichment of development related functions (Zhang et al., 2007a).

Up to now it has not been demonstrated that all H3K27me3 targets are targeted by PcG proteins, but a strong reduction of H3K27me3 on selected target loci has been observed in PcG mutants (Schubert et al., 2006). Thus, assuming that H3K27me3 marking is performed by PcG proteins, even though by different complexes, the terms “PcG target”

and “H3K27me3 target” will be used as synonym for genes carrying the histone mark in the following chapters.

2 Aim of study

The present study aims to identify and characterise novel shoot development related genes among H3K27me3 target genes. The underlying rational was based on the previous observation, that tissue specificly expressed sub sets of H3K27me3 target genes have a high probability to be involved in development and that there are many yet uncharacterised genes among those sub sets. It was postulated that there could be new genes involved in development related processes among the unknown genes of certain PcG target gene sub sets. To test this hypothesis, a group of shoot apex and floral organ expressed PcG target genes was screened for a role in shoot development.

This included, as a first part of the study, a transcriptional clustering analysis to obtain a cluster of apex expressed genes among H3K27me3 targets followed by Gene Ontology analysis to determine whether the selected cluster showed an overrepresentation of development related GO terms. Genes with unknown function in development among this cluster were screened for a role in shoot development by analysing either T-DNA insertion lines available in community stock centers or overexpressor lines created during this study. A screening scheme was established to identify putative developmental defects by monitoring several developmental traits in different conditions. Further analysis was performed to confirm the connection between the observed phenotypes and the analysed genes.

In a second part, this study focused on the detailed analysis of one candidate for which enhanced leaf serrations and enlarged petals were observed in loss-of-function mutants. The developmental function of this gene was further characterised via morphological and transcriptional studies.

Part I

Selection and Evaluation of candidate genes

Large gene lists such as the H3K27me3 target genes, can not be analysed on a single gene level. To extract certain sub sets of special biological interest and to analyse their characteristics, bioinformatic tools can be employed.

3.1 Clustering analysis of gene expression data

Clustering of gene expression data allows a subdivision of gene lists into groups of similar expression pattern. Any available expression data set, consisting of the genes of interest and their expression values in certain conditions, can be employed for this purpose. To obtain a group of genes with a certain expression characteristic, e.g. expression in a certain tissue, expression data over several tissues can be first clustered hierarchically to achieve an overview of the number of groups contained in the set, followed by k-means clustering to receive distinct gene list.

3.1.1 Hierarchical Clustering

Hierarchical clustering results are displayed as a tree where genes are assigned to branches and the length of branches reflects the distance between connected genes. Analysis of the cluster tree allows the identification of major branches (major expression patterns in the expression set), but the interpretation becomes complex with several hundreds or thousands of genes to be clustered. Therefore, in case of large data sets it is reasonable to use hierarchical clustering only on randomly selected sub sets of data with the aim to identify major branches present in the expression set. The number of major branches defines the number of clusters to be used in subsequent k-means clustering of the entire set (Eisen et al., 1998; Sturn, 2001).

3.1.2 K-means clustering

K-means clustering (Hartigan and Wong, 1979) is a method that groups genes together according to their expression pattern or any other information provided for the genes.

The number of groups (k) has to be specified in advance (normally the number of major branches determined by hierarchical clustering). Since k-means clustering leads to a defined number of gene groups, delimited lists of genes are generated. K-means clustering returns an average expression pattern for each cluster, which is called the center of the cluster. In an iterative process, the algorithm searches for both the best position for centers and the best assignment of genes to centers. First, cluster centers are randomly set and genes are assigned to the center with the lowest distance. Then the average for each cluster is calculated, and the center is relocated to this average value. Afterwards, the genes are assigned to the cluster center with the lowest distance. The averaging and reattribution is repeated for a number of iterations specified in advance, or until no further improvement in the overall distance of genes to their cluster centers can be achieved.

A disadvantage of the k-means algorithm lies in the random choice of the first cluster centers and its iterative character, which implies that there is no unique solution. Therefore, some genes are assigned to different clusters in two consecutive calculations based on the same expression data set and the calculation has to be repeated several times to determine those genes that are stably assigned to a given cluster of interest for further analysis.

3.1.3 Clustering of patterns or absolute values

If the aim of clustering analysis is to find genes similar in their expression patterns and their expression levels, absolute values can be clustered directly. The algorithm will use expression values as measured in the microarray experiment to calculate the euclidean distance between genes. This signifies that two genes, which are expressed in the same tissues/conditions but at very different levels, are located at a substantial distance and are not grouped together. Therefore, if the aim is to find genes with similar expression patterns regardless of their expression level, either expression values have to be normalised before clustering or a different distance measurement has to be employed. One normalisation possibility is to divide genes by their root mean square (rms) value. Division by rms (the mean for positive values) generates similar expression values between genes but preserves differences in pattern. Some clustering tools employs correlation coefficients (*e.g.* Pearson's correlation coefficient) to cluster genes according to expression patterns

regardless of their expression level (Sturn et al., 2002). In this case the product of the deviances from the mean divided by the product of the standard deviations for two values is used for the comparison.

3.2 Gene Ontology

The Gene Ontology (GO) is a general species-independent vocabulary that was built to standardise descriptions in three functional categories, which can be used to characterise a gene product. The GO categories are "biological processes", "molecular functions", and "cellular components". GO terms describe properties within these three functional categories. For example, a "biological process" can be a "developmental process", which is then further divided into 53 processes among which are "reproductive developmental process", "anatomical structure development", and "multicellular organismal development". The general term "biological processes" generates up to six sublevels in the GO vocabulary (the same holds true for "molecular function"; there are even more for "cellular component"). Accordingly, the GO is divided into levels. High levels correspond to general terms, and low levels to specific terms. The lower level terms are also called "child terms" of a general "parent term" to which they belong; *e.g.*, "developmental process" is a parent term of "anatomical structure development" and a child of "biological process". A GO child term can have several parent terms, thus creating a direct acyclic graph structure.

GO annotation databases contain gene products of an organism and their assigned GO terms. If a low level term is assigned to a gene product, all parents of the low level term are also automatically attributed to this gene. The GO annotation for Arabidopsis is curated by "The Arabidopsis Information Resources" (TAIR) database.

The attribution of GO terms to genes is based on different evidence, and, in consequence, not all GO term attributions are equally trustworthy. Usually, experimental evidence underlies the annotation, but there are also instances where the attribution of GO terms is purely based on *in silico* analysis. The method of investigation underlying the annotation of a GO term is indicated by an evidence code. For example, "IDA" stands for "Inferred from Direct Assay" and "ISS" for "Inferred from Sequence or Structural Similarity."

The GO annotation of an organism can be queried for single genes so that all GO terms are retrieved. In addition, the GO annotation can be queried for groups of genes, which allows to generate an overview about functions present in a large data set. In particular, GO slim terms were established for gene list queries. In the GO slim vocabulary, GO terms of lower levels are joined to one higher, more general level.

The advantage of the use of GO terms over keywords is that there is only one unique GO term for a function that may be defined by several keywords. The nonambiguity of GO terms is a precondition for statistical approaches that allow comparisons between different gene lists. These comparisons are necessary to determine whether specific functions are over- or underrepresented in a list of genes compared to a reference list (*e.g.* the whole genome). Several web suits offer tools to perform such functional enrichment analysis online (Rivals et al., 2007; Coulibaly and Page, 2008; Khatri and Drăghici, 2005). These tools calculate false discovery rate (FDR) corrected p-value for each GO term, which can be interpreted as the probability of this GO term to be not significantly overrepresented in the list submitted. Therefore, very low p-values correspond to very reliably overrepresented terms.

4.1 PcG target gene data sets

Two published data sets were used for preliminary experiments and comparison with internal data sets: genome-wide distribution data for the histone mark H3K27me3 published by Zhang et al. (2007a) and distribution data of LHP1 for the 4th Chromosome of Arabidopsis, published by Turck et al. (2007). For the main analysis, a list of LHP1 target genes obtained in our group prior to the beginning of this study was used. This list was obtained by Chromatin immunoprecipitation followed by hybridisation to whole genome tilling arrays (ChIP-chip). Used material were 10 day old seedlings of a *35S::LHP1:HA* line in Landsberg *errecta* (L*er*) ecotype grown on germination medium (GM) (0.5 strength Murashige and Skoog medium supplemented with 1% sucrose) under long day conditions. Material was immunoprecipitated with antibody against HA (αHA), followed by linker mediated amplification and hybridised to Roche-NimbleGen two colour array (ChIP sample and input sample were labelled with different colours) (Reimer and Turck, 2010). The following analysis was done with ChIPR (Göbel et al., 2010). LHP1 target genes were determined with the Rank Intersection Method implemented in ChIPR. The resulting list of LHP1 target genes contains 5057 genes and will be referred to as "H3K27me3 target genes" or "PcG target genes" in the following report, assuming co-localisation of LHP1 with the PcG related Histone mark H3K27me3 as shown by Turck et al. (2007) and Zhang et al. (2007b). The list of LHP1 target genes showed an overlap of 72% when compared to the already published data set for H3K27me3 targets, which contains 4979 genes (Zhang et al., 2007a).

Genome wide distribution for H3K27me3 was also determined for L*er* and Col-0 ecotype in the group, but after the beginning of this work. These lists were generated using ChIPR as well, but RINGO was used to determine enriched regions, which were subsequently mapped back to genes. These later generated gene lists were used for comparisons with the list used for candidate gene selection in this study and to display H3K27me3 distribution over single gene loci.

4.2 Expression data

For Arabidopsis numerous expression arrays have been analysed up to now and many of the results can be downloaded from the world wide web. In this project, the sub set "Developmental Series" of a data set called AtGenExpress created by Schmid et al. (2005) was used. This sub set contains expression data for different tissues in different developmental stages (79 samples in total, for detailed sample list see Appendix A, Tab. A.1). Of the 79 samples, 16 were obtain from plants with mutant backgrounds. Since this study aims to identify apex/floral organ expressed genes, the mutants used to generate apex (inflorescence apex) and flower (stage 12) samples are most interesting and are therefore mentioned here: *clv3-7*, *lfy-12*, *ap1-15*, *ap2-6*, *ap3-6*, *ag-12* and *ufo-1* (Wisman et al., 1998; Huala and Sussex, 1992; Ng and Yanofsky, 2001; Kunst et al., 1989; Yi and Jack, 1998; Hepworth et al., 2006). These mutants were mainly included in the series due to their different content of certain tissue or cell types: enlarged SAM (*clv3-7*), higher number of inflorescence branches (*lfy-12*, *ufo-1*) and homeotic transformations/ lack of certain floral organs (*ap1-15*, *ap2-6*, *ap3-6*, *ag-12* and *ufo-1*) (Clark et al., 1995; Schultz and Haughn, 1991; Mandel et al., 1992).

As a platform the AtGenExpress project used the Affymetrix®ATH1 array, each experiment was done as a triplicate. Expression sets were obtained from The Nottingham Arabisdopsis Stock Centre (NASC)(MAS 5.0 normalised data) and the group of Prof. Detlef Weigel (www.weigelworld.org)(gcRMA normalised data).

The Affymetrix®Microarray Analysis Suite 5.0 (Mas 5.0) performs a trimmed mean calculation: the mean of all values on the array, except the upper and lower 2% of values, is calculated, and than used to scale the data in a way that the mean value is equal to 100.

The gcRMA (robust multi-array average) algorithm (Irizarry et al., 2003) uses a global background and the GC-content information of the probes for correction. Expression values are calculated using a log_2 scale.

For each triplicate in the downloaded data sets, the mean value was calculated and samples were sorted according to contained tissue prior to further analysis (for list of samples in a certain tissue fraction see Appendix A, Tab. A.2).

4.3 Shannon entropy analysis

Shannon entropy was employed to determine tissue specificity as described in Schug et al. (2005). MAS 5.0 normalised expression data of the "Developmental Series" was used because of its linear values. All calculation steps were implemented in Perl programming language. According to Schug et al. (2005) relative expression values ($p_{t|g}$) for a gene (g) in a tissue/stage (t) were calculated as follows:

$$p_{t|g} = \frac{w_{g,t}}{\sum_{1 \leq t \leq N} w_{g,t}} \tag{4.1}$$

were $w_{g,t}$ is the expression value of a gene in a given tissue/ stage and N represents the number of tissues/ stages. From these relative expression values, the entropy (H_g) of a gene over all tissues/stages was calculated as:

$$H_g = \sum_{1 \leq t \leq N} -p_{t|g} log_2(p_{t|g}) \tag{4.2}$$

Entropy values are expressed in bits and they are zero for a gene expressed in a single tissue/ stage. The highest possible value of a gene expressed uniformly in all tissue is $log_2(N)$, which is $log_2(79) \approx 6.3$ for the "Developmental Series" data set.

4.4 Software and websuits

All clustering calculations were done with the software genesis (Sturn et al., 2002) as described in Engelhorn and Turck (2010). Genes were divided rms values over all expression values of one gene wherever patterns were to be observed. Functional enrichment analysis was performed using the websuit "babelomics" (Al-Shahrour et al., 2006) and its function FatiGO (Al-Shahrour et al., 2007). For versions 3 and higher the function "never remove duplicates" was enabled to prevent the program from removing genes from the lists. GO-Slim data was received from TAIR using "Functional categorisation" in the Gene Ontology Search field. To identify genes with high probability of redundancy, similar genes to each candidate were identified using wuBLAST results from the functional analysis tool AFAWE (Jöcker et al., 2008).

4.5 Plant material

All sequence-indexed Arabidopsis T-DNA insertion mutants were ordered from The Nottingham Arabisdopsis Stock Centre (NASC). They were either produced by the Salk Institute Genomic Analysis Laboratory (SIGnAL), the Syngenta Arabidopsis Insertion Library (SAIL), the WiscDsLox collection or the Gabi-Kat consortium (Sessions et al., 2002; Rosso et al., 2003; Woody et al., 2007) (for complete list of the T-DNA insertion lines used in this study see Appendix A, Tab. A.3). As wildtype control Col-0 plants were used, since this is the background of all four collections. Col-0 ecotype was as well used for generation of transgenic plants.

4.6 Growth conditions

If not otherwise indicated, experiments were conducted with soil grown plants. For long days (LD) of 16 h light and 8 h darkness plants were grown in greenhouse at 20°C, for short days (SD) of 8 h light and 16 h darkness plants were either grown in growth cabinets at 20°C, 70% humidity or in greenhouse (just for confirmation of phenotypes observed already in cabinet). Prior to sowing, seeds were stratified at 4°C for 4 days on wet paper. Ages indicated are days after sowing. Seeds of lines that did not germinate on soil were sown on GM.

4.7 DNA extraction and purification

4.7.1 DNA extraction from plant material

Low amounts of genomic DNA from plant material (usually one young leaf) were extracted with the BioSprint 96 robotic work station (Qiagen) using the program DNA-plant-100. For higher amounts, DNA was extracted using Qiagen DNeasy®kit. Reagents were used according to manufacturer's instructions.

4.7.2 DNA extraction and purification from bacteria

Plasmids were purified using Nucleospin®Plasmid kit (Machery-Nagel), fragments from agarose gels were purified using Nucleospin®Extract II kit (Machery-Nagel) according to manufacturer's instructions.

4.8 Genotyping of T-DNA insertion lines

Genotyping was performed using primer pairs suggested by the webtool "T-DNA Primer Design" on the Salk Institute website (signal.salk.edu/tdnaprimers.2.html). To detect the wild type allele, two primers for each insertion line were ordered, flanking the indicated insertion. One primer was designed to be close to the left border of the T-DNA (LP primer) and one about 1000 bp away from the predicted right border (RP primer) (for a list of all primers see Appendix A, Tab. A.3). To detect the allele containing the T-DNA insertion, RP primer and a primer complementary to a part close to the left border on the T-DNA was used (LB primer). LB primers for the four T-DNA insertion types were:

Salk-lines:

LBb1.3 - 5´-ATTTTGCCGATTTCGGAAC-3´

SAIL-lines:

LB2-short - 5´-GCTTCCTATTATATCTTCCCAAATTACC-3´

WiscDSLox-lines:

LB-WisDscLox- 5´-AACGTCCGCAATGTGTTATTAAGTTGTC-3´

Gabi-Kat-lines:

LB-GABI - 5´-ATATTGACCATCATACTCATTGC-3´

Polymerase chain reaction (PCR) was done separately for each primer pair (LP/RP and LB/RP) in 20 μl of the following reaction mix: 1x Reaction buffer (50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl pH 8.5, 0.01 % (w/v) gelatine, 0.05 % (v/v) Tween20, 0.0025 % (w/v) BSA) containing 1x Creosol red loading dye (1.5 % (w/v) sucrose,0.02 % (w/v) creosol red (Aldrich)), dNTPs (1.25 mM each), Primer-mix (0.5 mM each), Taq DNA Polymerase (0.3125 U/μl) and 3 μl template DNA.

Amplification was done using the following cycles:

94°C	2	min	
94°C	30	sec	35 cycles
54°C	30	sec	
72°C	1	min	
72°C	5	min	
15°C	∞		

Elongation step for LP/RP reaction was 15 sec longer. 10 μl of each reaction were mixed and loaded to an agarose gel for simultaneous analysis.

4.9 Screen for developmental defects

4.9.1 Flowering time

Flowering time was measured by scoring the number of rosette and cauline leaves on the main stem of 10 to 12 individuals. To determine the number of days until flowering, days from sowing until bolting were measured, bolting was defined as the time when about 1 cm of the main shoot was visible.

4.9.2 Rosette diameter

Rosette diameters were measured at time of bolting using a ruler. To account for asymmetries, rosette diameters were measured in different orientations and only the highest diameter observed was recorded.

4.9.3 Other parameters

Other developmental parameters were only recorded, if a difference compared to wild type could be observed by eye. Parameter checked in this context were:

- leaf shape
- leaf colour
- phyllotaxis (in terms of silique position)
- flower size
- petal number
- branching

4.10 Generation of plants overexpressing candidate genes

4.10.1 Construction of vectors

To generate overexpression lines, the coding regions of the representative gene model of candidate genes were amplified from Arabidopsis (Col-0) cDNA using specific primers with GATEWAY™ extensions (primers are listed in Appendix A, Tab. A.4). The forward primers contained the attB1 extension (5´-GGGGACAAGTTTGTACAAAAAAGCAGG-CTCC-3´); reverse primers contained the attB2 tail (5´-GGGGACCACTTTGTACAAG-AAAGCTGGGTC-3´). Both sequences contained also linker bases to enable in-frame cloning of the amplified product into a GATEWAY™ entry-vector. Amplified fragments were cloned into pDONR201 or pDONR207 (Invitrogen) via BP reaction. BP products were sequenced to eliminate PCR errors. Entry clones containing the correct candidate gene sequence were introduced into the binary pAlligator2 vector (Bensmihen et al., 2004) by LR reaction. BP and LR reactions were performed according to manufacturer's instructions but with one fourth of the recommended reaction volume and an overnight incubation instead of one hour. pAlligator2 allows overexpression of the inserted sequence via the cauliflower mosaic virus 35S promoter. In addition to that, the coding sequence is translationally fused to the coding sequence of a triple HA (hemagglutinin)

tag. This leads to an N-terminal fusion of 3 x HA to the protein of interest which later on allows detection and purification. All vectors were propagated in *Escherischia coli* (*E.coli*) DH5α strain (for vectors containing candidate gene sequences) and DB3.1 strain for empty vectors.

4.10.2 Generation of transgenic plants

pAlligator vectors containing the candidate gene sequences were transformed into *Agrobacterium tumefaciens* ASE strain (Fraley et al., 1985) by electroporation. Arabidopsis plants (Col-0) were transformed with the vectors via the floral dip method (Clough and Bent, 1998).

4.10.3 Selection of transformants and generation of homozygous lines

pAlligator vectors contain an endoplasmic reticulum targeted GFP (E-GFP) coding sequence driven by the seed coat specific Arabidopsis *SEED STORAGE ALBUMIN* (*3At2S3*) promoter (Bensmihen et al., 2004). This allows selection of transgenic seeds via the GFP signal. Selection was done using Leica MZ16 FA stereomicroscope with GFP3 filters (excitation: 450 – 490 nm, barrier: 500 – 550 nm). T_2 generations were tested for single locus insertion events by segregation ratio. Only lines that showed fluorescent seed to non-fluorescent seed ratios between 2:1 and 4:1 were selected for further analysis.

4.11 Expression analysis

Plants grown on soil were cut above the soil, for seedlings/ calli grown on plates, whole plants were harvested. Samples named “flowers” contained flowers from stage 14, “apex/ inflorescence” (apex/infl.) samples contained main inflorescences including closed flowers. Samples specified as “apex” were harvested from vegetatively growing plants. In this case all leaves larger than 5 mm were removed, as well as the roots. About 2 – 5 mm of hypocotyl remained in the sample. For “leaf” samples, the leaves removed for “apex” samples were collected. This results in apex or leaf enriched samples, but samples did not contain exclusively the respective tissue. Harvested material was immediately frozen in liquid nitrogen and stored at −80°C until further processing. Total RNA was isolated us-

ing RNeasy®kit (Qiagen), 5 μg of RNA were DNase treated with DNA-*free*™ kit (Ambion). cDNA synthesis with oligo-dT priming was performed with SuperScript™ II reverse transcriptase (Invitrogen) and dT18 primer according to suppliers instructions. Final cDNA reactions were diluted to 150 μl with water, 3 μl of diluted cDNA were used for quantitative real time polymerase chain reaction (qRT-PCR) using BioRad iCycler iQ5™ Real-Time PCR Detection System with EvaGreen® dye (Biotium) to detect the product. Final reaction volume was 20 μl (10 μl 2x EvaGreen® buffer, 2 μl Primer-mix (10 mM each), 3 μl cDNA, 5 μl H_2O).

2x EvaGreen® buffer (for 100 reactions):

10x Reaction buffer (0.8 M KCl, 50 mM $MgCl_2$ 0.2 M, Tris-HCI pH 8.8)	100 μl
EvaGreen®	100 μl
dNTPs (10 mM each)	40 μl
Taq DNA Polymerase (5 U/μl)	20 μl
H_2O (sterile)	740 μl

Amplification was performed using a two-step cycling program:

95°C	3	min	
95°C	10	sec	} 45 cycles
60°C	45	sec	

A dilution series of an “apex” sample was used as standard for each primer pair (for primer sequences see Appendix A, Tab. A.5). To account for variations in cDNA content between samples, data was normalised using an ubiquitously expressed gene, encoding for the regulatory subunit of protein phophatase 2A (PP2A, At1g13320) (Martin-Tryon et al., 2007).

5.1 Entropy analysis on H3K27me3 target genes

To investigate whether the data set of H3K27me3 targets used in this study contains a higher fraction of tissue specific genes than the whole genome (as does the already publish data set (Zhang et al., 2007a)), a Shannon entropy analysis was performed (Fig. 5.1) using the "Developmental series" expression data set (Schmid et al., 2005). The AtGenExpress data set was generated using the ATH1 expression array, therefore not all genes are present on the array. In the Mas 5.0 normalised data set used for the entropy analysis values exist for 3346 genes in the H3K27me3 target list, 3417 genes in the published data set and 22769 genes in total. For the 79 samples of the "Developmental series" an entropy value of 6.3 bits is expected for a gene equally expressed in all tissues and developmental stages and 0 bits for a gene expressed only in a single tissue at a single stage. Such extreme values were not observed among this data set, indicating that all genes are at least expressed in several tissues or at several developmental stages in one tissue and that there is at least some variation in the expression level for each gene. The lowest observed entropy value was 0.41 bits, the highest was 3.99 bits.

For both H3K27me3 target lists the distribution of gene fractions over the entropy values is shifted to lower entropy values compared to the list containing all genes, meaning that the H3K27me3 target lists contain higher fraction of tissue specific genes and lower fraction of genes with rather equal distribution over tissues when compared to all genes. There is no peculiar difference between the two H3K27me3 target lists, which indicates that the data set used in this study is as tissue specific as the already published one. This observation was taken as an additional hint that the H3K27me3 data set used in this study is well suited for further analysis.

5.2 Clustering analysis

Clustering analysis was performed to identify a group of genes among the H3K27me3 target genes that is mainly expressed in the shoot apex, since major developmental pro-

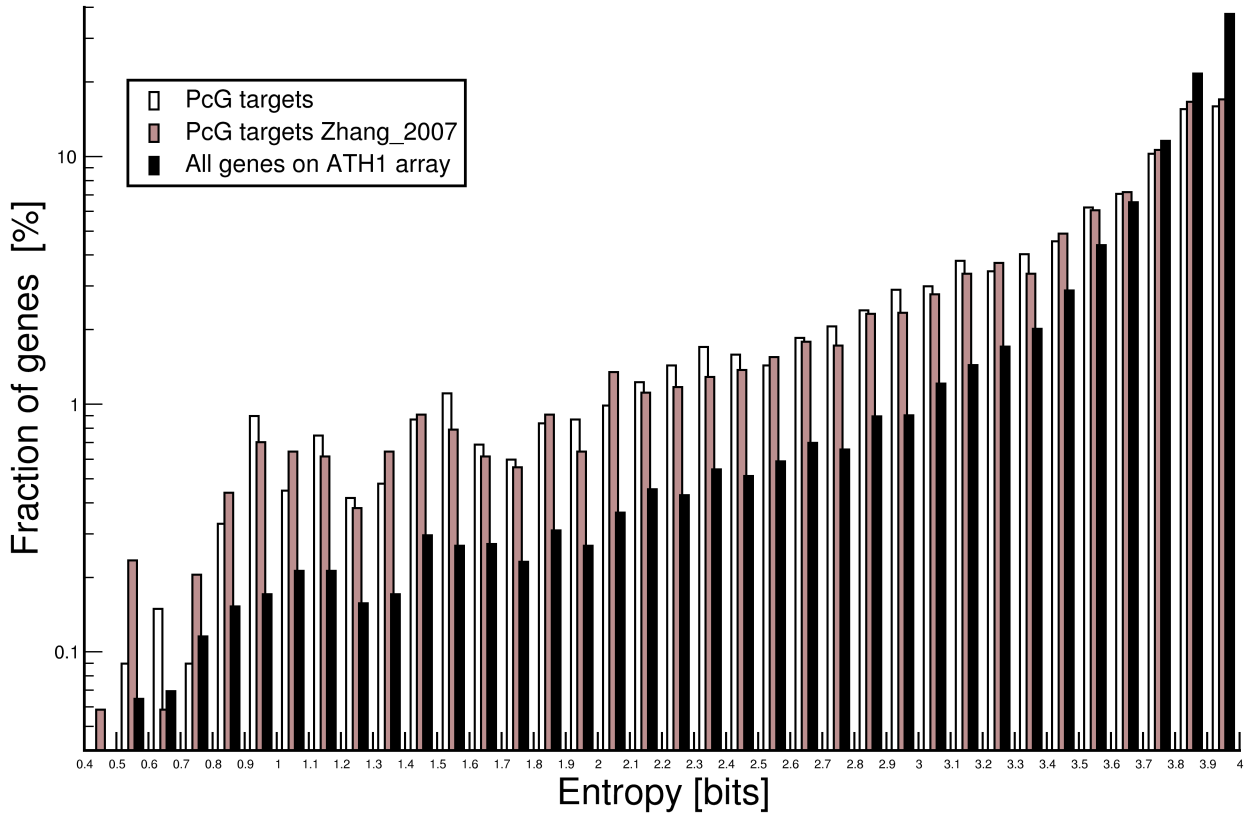


Figure 5.1: **Shannon entropy values for H3K27me3 targets compared to all genes.** Shannon entropy was calculated using the "Developmental series" expression data set. The analysis was performed for genes of the H3K27me3 target list used in this study, for genes in the previously published H3K27me3 target list (Zhang et al., 2007a) and all genes present in the Mas 5.0 normalised data set. Observed entropy values were divided into bins of 0.1 bit and fraction of genes for each data set were calculated. Values on x-axis indicate borders of bins. Y-axis is displayed in a logarithmic scale to emphasise bins with lower fractions of genes since there the most noticeable differences are observed.

cesses are controlled in this tissue. An overview of the work flow used to finally obtain candidates for a putative new role in development is displayed in Fig. 5.2.

5.2.1 Hierarchical clustering

For a first overview about the groups of expression patterns present in the H3K27me3 target gene set, hierarchical clustering was employed. Since a tree containing expression data for all H3K27me3 targets would be rather complicated to interpret (because of its length), a hierarchical clustering analysis of a sub set of 500 randomly chosen H3K27me3 target genes using the "Developmental series" data set was performed. Among the 500

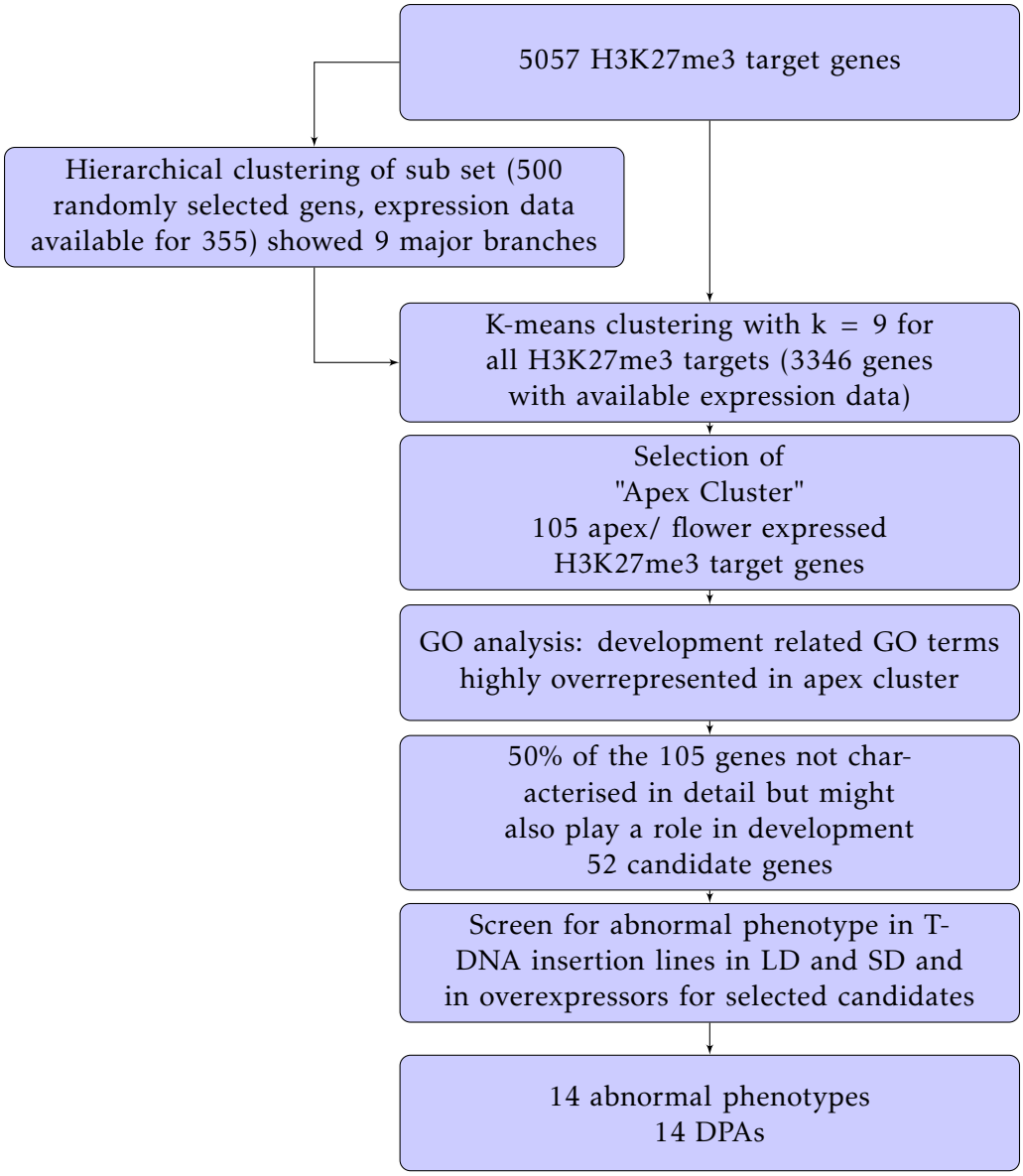


Figure 5.2: **Work flow for candidate selection.**

genes in the set, 355 genes were present on the ATH1 array and contributed to the analysis. This pre-analysis revealed nine major branches, indicating nine groups of genes with similar expression pattern (Fig. 5.3). Therefore, k = 9 was used for the following k-means clustering analysis.

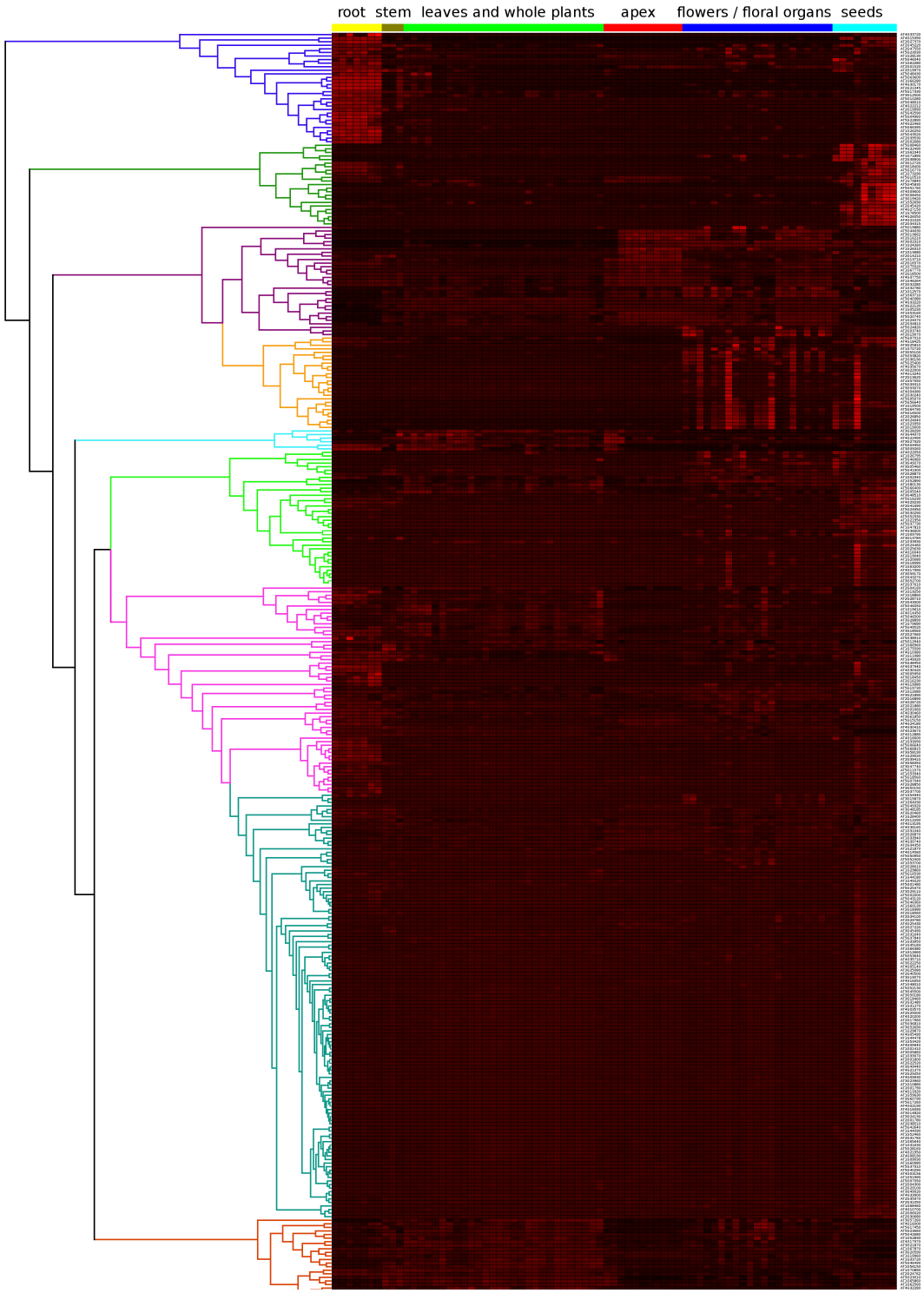


Figure 5.3: **Hierarchical clustering for 355 randomly chosen H3K27me3 target genes to determine number of major branches.** gcRMA normalised expression values were divided by rms to cluster according to expression pattern instead of absolute values. Expression values are indicated in a heat map, light red indicates highest expression, black indicates no expression. Major branches corresponding to major expression patterns are highlighted by different colours. The analysis was repeated several times with different random gene sets with similar results.

5.2.2 K-means clustering

K-means clustering with k = 9 was performed for genome-wide H3K27me3 targets with available expression data (3346 genes for Mas 5.0 data set and 3571 for gcRMA data set) and all genes present on the ATH1 expression array of Affymetrix®for both Mas 5.0 (22769 genes) and gcRMA (22810 genes). All clusters except one, which contains mainly lowly expressed genes, show a very tissue specific expression pattern (Fig. 5.4, example for gcRMA data set). One cluster (cluster 4 in this case) could always be identified as a group of mostly apex and flower/floral organ expressed genes (called "apex cluster" in the following sections). This cluster was then used for further analysis.

To assure that k = 9 was a reasonable number of clusters, the gcRMA normalised data set for H3K27me3 targets was also clustered with k = 8 and k = 10 (data not shown). When k = 8 was used, the apex cluster was merged with the cluster containing genes expressed higher in flowers than in other tissues (cluster 7 in Fig. 5.4). When k = 10 was used, the cluster containing mainly seed-specific genes (cluster 6 in Fig. 5.4) split up. This confirms k = 9 as a reasonable number of clusters to obtain genes with main expression in the shoot apex part of the plant from the "Developmental series" data set.

As mentioned above, one disadvantage of the k-means algorithm is its iterative character, which means, that there is no unique solution. Therefore, the calculation was repeated several times for each data set (Mas 5.0 and gcRMA). 10 repeats seemed to be sufficient for this purpose, since all genes common in 9 list were also all found in the 10^{th} list in the case of Mas 5.0. In the case of gcRMA data the core data set was more robust: the genes common in the first two lists were also found in all following ones. As result, 151 robustly apex expressed genes for Mas 5.0 data and 209 genes for gcRMA data were obtained. Comparison of both lists resulted in a list of 105 candidate genes which were always assigned to the apex cluster in both data sets.

The same analysis was performed for all genes present on the ATH1 array (which was used to generate the expression data). An apex cluster could also be obtained among all these genes. It was treated in the same way as the H3K27me3 apex cluster (10 repeats for both of the data sets) and resulted in 1229 apex expressed genes. This set was used for comparison in the GO analysis.

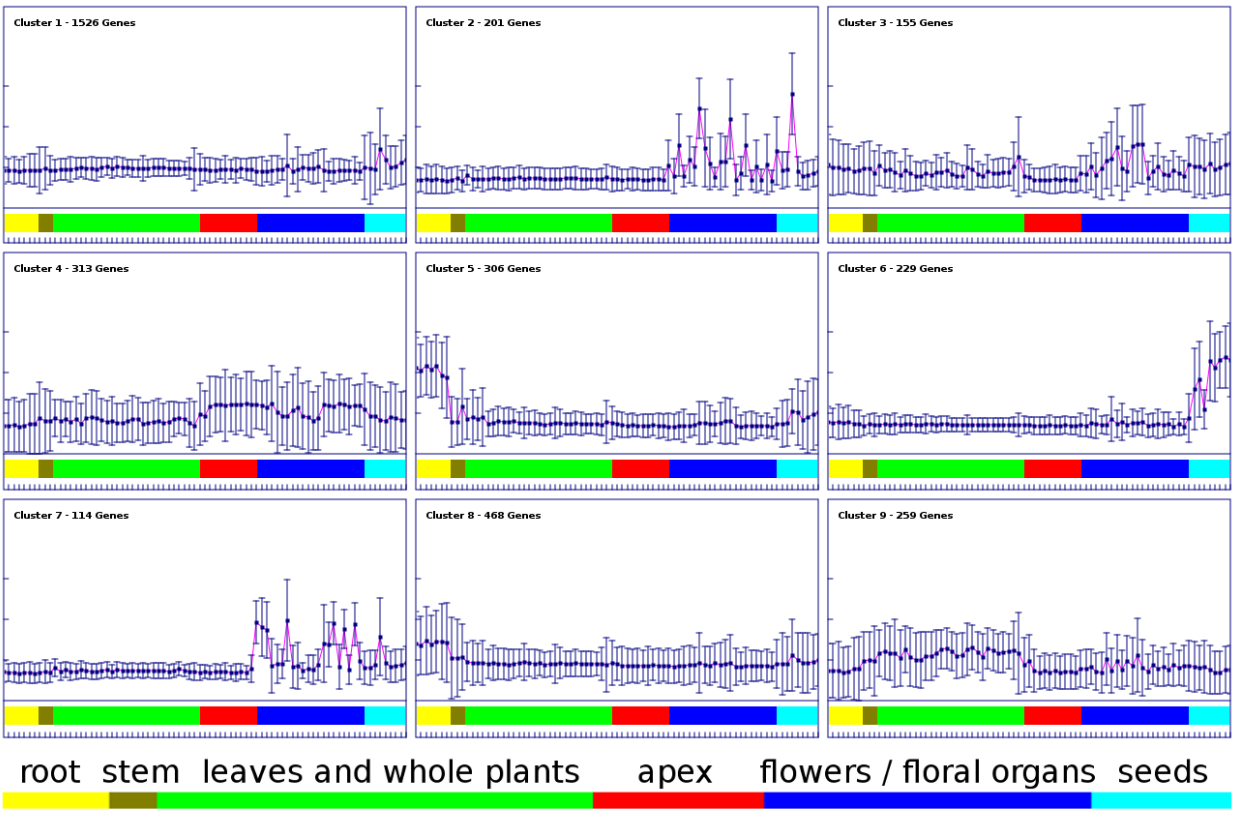


Figure 5.4: **K-means clustering of H3K27me3 target genes,** k = 9**.** Each of the boxes represents one cluster. Different tissues are presented on the x-Axis according to the colour code displayed at the bottom, expression values on the y-Axis. Every square represents the average expression of one tissue, bars indicate the variance within one tissue.

5.2.3 Characteristics of the apex cluster

Since the cluster analysis was performed with rms-normalised expression values, only the pattern was concerned for the selection of genes. Therefore, all genes in the cluster should be higher expressed in apical tissues than in other tissues, but their absolute expression level was not considered. Within the cluster, two types of expression pattern are distinguishable: several genes are lowly expressed in all other tissues except apex and flowers whereas the other genes are expressed in all tissues but higher in the apex and flower samples (Fig. 5.5). This observation raises the question, whether one of these groups might be a better suited subgroup for the discovery of novel genes with development functions and if the apex cluster should be further divided for the following analysis. A preliminary GO slim analysis (data not shown) estabished that genes from the first category (54 genes) display a higher percentage of genes with developmental GO

terms than genes from the second category (51 genes). As the latter were also more characterised, the chances were that more development related genes in the second category had not yet been discovered. Therefore, no further subdivision was performed and the complete apex cluster of 105 genes was used for further analysis.

According to previous observations, tissue specific expressed clusters of the PcG targets should contain a high number of developmentally relevant genes (Zhang et al., 2007a). Indeed, many important genes with known role in shoot development affecting traits like leaf and flower development (e.g. *AS2*, *JAGGED* (*JAG*)), auxin biosynthesis and transport (*YUC4*,*PIN6*), meristem initiation and organ boundary specification (*CUC2* and *CUC3*), meristem maintenance (e.g. *KNAT2*, *KNAT6*, *APETALA1* (*AP1*)) and embryonic and cotyledon development (e.g. *PROTODERMAL FACTOR2* (*PDF2*), *WUSCHEL-RELATED HOMEOBOX 9* (*WOX9*)) were found among the H3K27me3 apex cluster genes (see Appendix A, Tab. A.7 for the complete list).

5.3 Functional enrichment analysis

5.3.1 GO slim

The distribution of GO slim terms in the apex cluster gene list was compared to the distribution of GO slim terms in the complete genome (Fig. 5.6). This revealed that over 30 % of the apex cluster genes were involved in development at this point (time of analysis beginning of 2008), whereas only 5 % of all genes had a developmental function assigned. This result can be caused by three properties of these genes: either by the apex expression, by being a H3K27me3 target or by a combination of both. To test which property causes the overrepresentation, the analysis was also done for all H3K27me3 targets and for all apex expressed genes derived from the cluster analysis. In neither of the two additional lists a percentage as high as for the apex cluster H3K27me3 targets could be found, suggesting that the enrichment of genes involved in developmental processes in the apex cluster gene list is not caused by the apex expression or by being a H3K27me3 per se target but by being both.

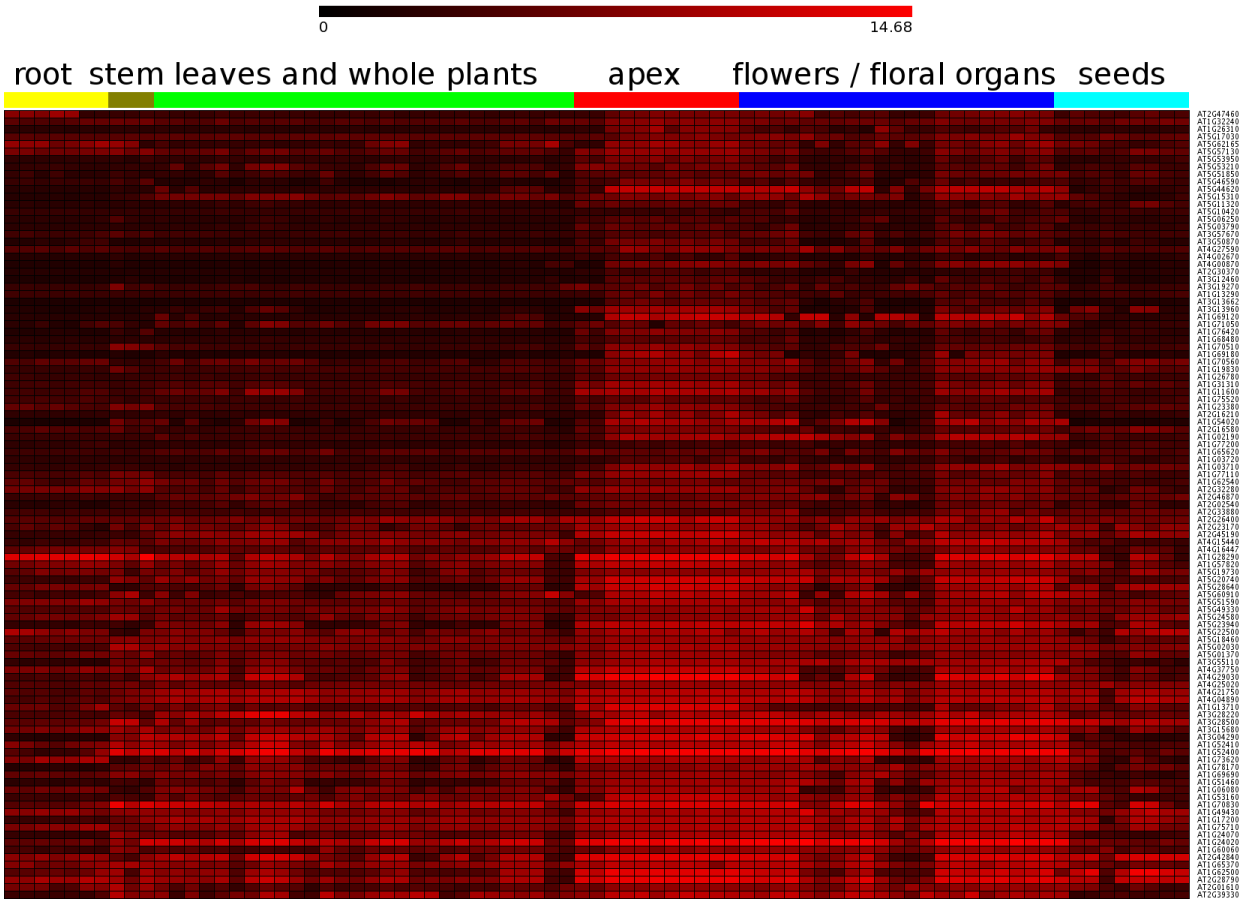


Figure 5.5: **Heat map of expression in apex cluster genes.** Absolute expression values (gcRMA normalised data, log_2 scale) are indicated by a colour code, light red indicates highest expression occurring in the data set, black indicates no expression.

5.3.2 GO full

Observations based on GO slim analysis were confirmed by a functional enrichment analysis using the GO full annotation in the web suite FatiGO (Al-Shahrour et al., 2007). This tool returns all significantly overrepresented GO terms in a submitted list compared to a reference list with a FDR-corrected p-value for each term. As a reference list, all genes present on the ATH1 array (only genes present in Mas 5.0 and gcRMA data set, 21258 genes) were used, since these genes had a potential to be in the candidate list. As for the previously tested groups of H3K27me3 targets expressed in the same tissue types (Zhang et al., 2007a), several developmental GO terms were found to be overrepresented among the apex cluster list. Again this overrepresentation could be caused by the apex

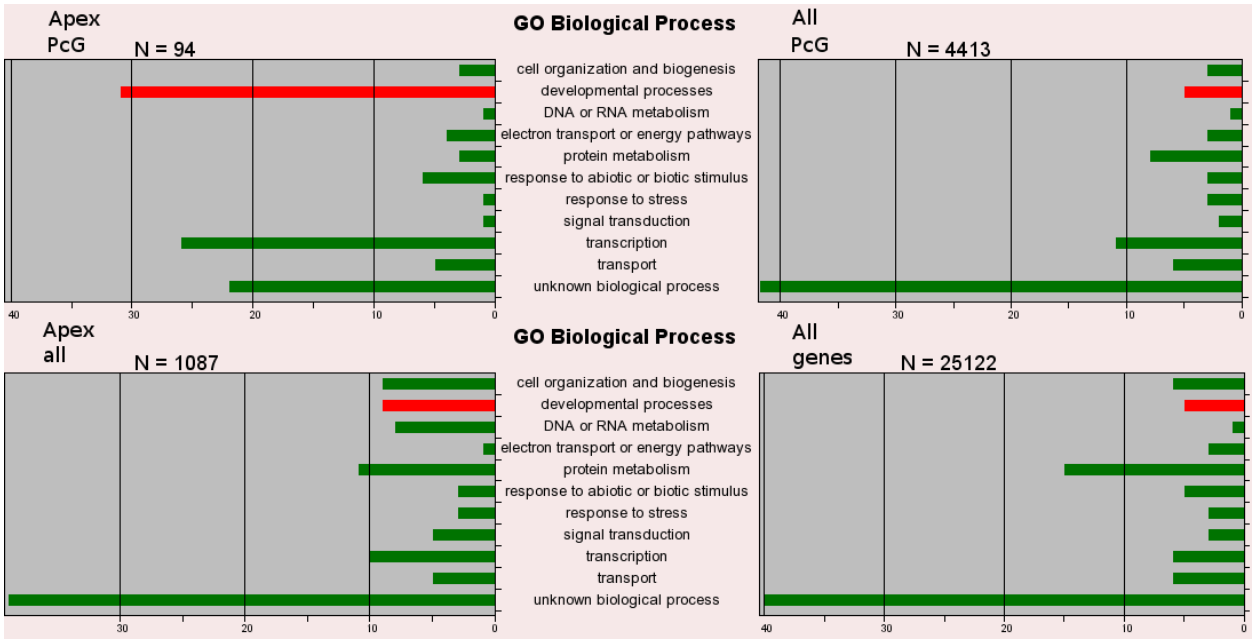


Figure 5.6: **Distribution of GO slim terms in H3K27me3 apex cluster, all H3K27me3 targets, all apex expressed genes and all genes.** Output of the "Functional categorisation function in the Gene Ontology Search field of TAIR. X-axis indicates fraction of genes which were assigned the respective GO term in percentage. Bars for the term "developmental process" are highlighted in red.

expression or by the fact of being a H3K27me3 target alone. To rule this out, the same analysis was also done for all H3K27me3 targets present on the ATH1 array and for all apex expressed genes derived from the cluster analysis. Several development related GO terms were found to be only overrepresented for the apex cluster of the H3K27me3 targets (Tab. 5.1). The others (except anatomical structure formation) show at least one magnitude difference in the *p*-values. This result confirmed the suggestion that being a H3K27me3 target and being expressed in the apex makes it more probable for a gene to play a role in development and led to the hypothesis that the unknown genes among the apex cluster list might also play a role in development.

5.4 Analysis of candidate genes

The genes contained in the apex cluster were submitted to the search function of the "The Arabidopsis Information Resource" (www.arabidopsis.org) to receive information about GO terms, available publications and characterised mutants. Only genes which had ex-

Table 5.1: ***P*-values of overrepresented GO-terms for PcG apex cluster, all PcG targets and all genes.** Overrepresented GO-terms only found for apex expressed PcG targets are printed bold.

GO-Term	Apex PcG	Apex all	PcG all
anatomical structure development (GO:0048856)	$1.09 \cdot 10^{-10}$	$5.41 \cdot 10^{-6}$	$1.26 \cdot 10^{-6}$
anatomical structure formation (GO:0048646)	$2.05 \cdot 10^{-2}$	$8.14 \cdot 10^{-4}$	
anatomical structure morphogenesis (GO:0009653)	$5.14 \cdot 10^{-8}$	$4.93 \cdot 10^{-4}$	
axis specification (GO:0009798)	$8.50 \cdot 10^{-5}$		
carpel development (GO:0048440)	$2.43 \cdot 10^{-2}$		
DNA binding (GO:0003677)	$2.58 \cdot 10^{-11}$	$1.08 \cdot 10^{-2}$	
floral organ development (GO:0048437)	$1.36 \cdot 10^{-2}$		
floral whorl development (GO:0048438)	$1.21 \cdot 10^{-3}$	$1.70 \cdot 10^{-2}$	
flower development (GO:0009908)	$9.52 \cdot 10^{-7}$	$8.13 \cdot 10^{-3}$	$1.57 \cdot 10^{-3}$
fruit development (GO:0010154)	$1.92 \cdot 10^{-3}$		
gynoecium development (GO:0048467)	$2.48 \cdot 10^{-3}$	$2.43 \cdot 10^{-2}$	
leaf development (GO:0048366)	$3.11 \cdot 10^{-2}$		
meristem development (GO:0048507)	$1.70 \cdot 10^{-3}$		
multicellular organismal development (GO:0007275)	$1.09 \cdot 10^{-10}$	$5.66 \cdot 10^{-8}$	$6.20 \cdot 10^{-7}$
nucleic acid binding (GO:0003676)	$1.71 \cdot 10^{-7}$	$3.43 \cdot 10^{-2}$	
organ development (GO:0048513)	$7.66 \cdot 10^{-10}$	$7.97 \cdot 10^{-6}$	$2.79 \cdot 10^{-4}$
organization of an anatomical structure (GO:0048532)	$2.33 \cdot 10^{-2}$		
pattern specification process (GO:0007389)	$9.76 \cdot 10^{-7}$	$1.10 \cdot 10^{-3}$	$4.27 \cdot 10^{-2}$
phyllome development (GO:0048827)	$1.36 \cdot 10^{-2}$		
positive regulation of developmental process (GO:0051094)	$2.38 \cdot 10^{-2}$		
post-embryonic development (GO:0009791)	$4.88 \cdot 10^{-7}$	$6.51 \cdot 10^{-3}$	$3.55 \cdot 10^{-3}$
post-embryonic organ development (GO:0048569)	$6.73 \cdot 10^{-3}$		
regulation of developmental process (GO:0050793)	$6.25 \cdot 10^{-4}$		
regulation of transcription (GO:0045449)	$8.75 \cdot 10^{-4}$		
reproductive developmental process (GO:0003006)	$5.11 \cdot 10^{-6}$	$8.71 \cdot 10^{-4}$	$7.57 \cdot 10^{-4}$
reproductive structure development (GO:0048608)	$2.20 \cdot 10^{-5}$	$8.14 \cdot 10^{-4}$	$5.09 \cdot 10^{-4}$
shoot development (GO:0048367)	$4.15 \cdot 10^{-3}$	$4.31 \cdot 10^{-2}$	
shoot system development (GO:0022621)	$2.59 \cdot 10^{-3}$	$3.15 \cdot 10^{-2}$	
specification of organ axis polarity (GO:0010084)	$1.15 \cdot 10^{-3}$		
system development (GO:0048731)	$3.80 \cdot 10^{-10}$	$2.20 \cdot 10^{-5}$	$1.72 \cdot 10^{-4}$
tissue development (GO:0009888)	$1.08 \cdot 10^{-3}$	$2.83 \cdot 10^{-2}$	
transcription (GO:0006350)	$3.26 \cdot 10^{-5}$		
transcription factor activity (GO:0003700)	$2.02 \cdot 10^{-14}$		
transcriptional activator activity (GO:0016563)	$8.96 \cdot 10^{-3}$		

clusively very general GO terms and publications assigned and of which no mutants were characterised were considered for further analysis. 52 of the candidate genes belonged to that category and were included in the screen for an abnormal developmental phenotype described in the following sections (see Appendix A, Tab. A.7 for a complete list of candidate genes). The first strategy was to obtain T-DNA insertion lines in the candidate gene loci to screen for abnormal phenotypes in putative knock-down lines for each candidate. For genes with putative redundancies, genes where no T-DNA lines were available and some genes associated with abnormal phenotypes in T-DNA insertion lines, overexpressor lines were created (genes and screened conditions are summarized in Tab. 5.2 and 5.3). Genes, for which abnormal phenotypes concerning development were observed were considered for further analysis and named putative Development related PcG Targets in the Apex (DPAs).

5.4.1 Analysis of T-DNA insertion lines

T-DNA insertion lines were ordered for 50 candidates (Appendix A, Tab. A.3). Only for two, At4g16447 and At3g15680, no lines are available. In cases where only promoter or intron insertions were available, more lines where ordered, if possible. An internal code consisting of a letter and a number (e.g. A1) was given to each T-DNA insertion line for faster recognition. In three cases, no T-DNA could be detected by PCR in the available lines (At5g20740 (Line A9), At1g28290 (Line C9) and At2g28790 (Line D9)). For At5g20740 and At2g28790 more lines are available now and will be characterised in future. One line did not germinate on soil (At5g17030 (Line A8)) but recently one plant was recovered from a GM plate and will be analysed in future.

Therefore, 45 candidates with available, genotypable T-DNA insertion lines were left for the screen. These lines were first screen in LD conditions, since plants grow fastest in this condition. Four lines with abnormal phenotypes could be identified in this screen: a late flowering line, one with small, bushy plants, a line with smaller petals compared to wild type and another were no homozygous plants could be obtained (indicating that the insertion in this gene might be lethal when homozygous)(Tab. 5.4). T-DNA lines for which abnormal phenotypes could be observed were then further analysed to confirm the connection of the phenotype to the candidate gene (Section 5.5).

Arabidopsis development is much slower in SD and flowering occurs later. In SD, early flowering mutants which are independent of the photoperiodic pathway can be observed and defects in vegetative development can become visible due to the prolongation of this developmental phase. Since only a few abnormal phenotypes were observed in the LD screen, all candidates for which homozygous T-DNA insertion lines were available were also screened in SD. For At2g01610 no line was screened, since the only available line (A13) was received just recently. Eight additional lines with abnormal phenotypes were identified in this screen: one line displaying stronger leaf serrations and enlarged petals, three early flowering lines (one line with significantly reduced leaf number until bolting (Student's t-test, $p > 0.05$) and two lines which were not significantly early but did also not correspond to Col-0 values within the standard error, described as slightly early flowering in the following paragraphs), two lines with reduced rosette diameter, a line with reduced apical dominance and a line where 30 % of the plants display extremely small, dark leaves and did not flower in SD for 3 months (Tab. 5.4). Again, T-DNA lines for which growth alterations could be observed where then further analysed to confirm the connection of the phenotype to the candidate gene (Section 5.5).

5.4.2 Analysis of overexpressor lines

Overexpression lines for 21 candidate genes were created, 17 of them have been screened at least in T1 generation in LD conditions (strong phenotypes should be already visible in T1 generation, as GFP marker selection allows stress free selection without herbicides). The final goal of this part of the screen is to obtain overexpressors for all candidate genes. Since this could not be done in parallel for all candidates in one step, candidates were ordered by priority and then step by step cloned and transformed into plants. First overexpressors for genes with high sequence similarity to other genes were created. Five genes with more than 60 % identity to other genes were selected: At5g46590, At1g02190, At1g73620, At1g19830 and At5g51590. Among those overexpressors, only for the At5g51590 overexpressor an abnormal phenotype could be observed in several independent lines. Next, overexpressors were generated to confirm phenotypes observed in T-DNA insertion lines (7 cases: At1g75710, At1g69690, At5g18460, At5g06250, At1g62540, At1g78170 and At1g65370) and for the two candidates with no available T-DNA

insertion lines. The remaining 7 genes were selected on a random basis. No additional abnormal phenotypes were observed in those lines so far.

5.4.3 Candidates characterised by other research groups

For three of the 52 candidate genes, articles were published during this screen in which their abnormal phenotype was described. These were At1g26780, At1g13290 and At1g-70560. For At1g13290 and At1g70560, publications came out before these candidates were intensively screened and therefore no statement is possible, whether the abnormal phenotype would have been discovered in this screen as well. The At1g70560 gene product was published as TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS 1 (TAA1), an enzyme of the indole-3-pyruvic acid (IPA) branch of the auxin biosynthetic pathway (Stepanova et al., 2008). Single knock out mutants of *TAA1* are not reported to display any phenotype but a weak insensitivity to ethylene in root growth and a loss of gravity sensing potential. In contrast to that double mutants of *TAA1* and *TRYPTOPHAN AMINOTRANSFERASE RELATED 2* (*TAR2*) display decreased apical dominance and inflorescences with abnormal flowers with reduced number of organs caused by reduced levels of auxin. At1g13290 was characterised to be involved in vein patterning and was named DEFECTIVELY ORGANIZED TRIBUTARIES 5 (DOT5) (Petricka et al., 2008). *dot5* mutants display defects in phyllotaxis, delayed leaf initiation and reduced apical dominance. For At1g26780, which displays organ fusions and twisted side shoots when knocked out, the phenotype was also observed in this screen and the gene was included in the DPA list as *DPA6*. The gene product of At1g26780 encodes for LATERAL ORGAN FUSION1 (LOF1) and was published by Lee et al. (2009) and therefore excluded from the further screening process.

All three genes are involved in plant development, strongly supporting the assumption that there are several developmental genes among the candidates analysed in this study.

Table 5.2: **Summary of screening progress for candidate genes 1.** Only candidates with available T-DNA insertion or overexpressors (OE) are mentioned here. The ✓icon indicates that at least one line was assessed in the condition/ genotype. Sum displays number of genes for which the described condition/ genotype has been screened, numbers in brackets indicate number of lines evaluated.

Gene	Lines ordered	Homozygous lines	SD	OE T_0	OE T_1	OE T_2	OE T_3
At1g75710	C/D2, B5, C-H10, B12	C2, C/D10, F-H10	✓	✓	✓	✓	✓
At5g51590	I3, J3, A5	I3, J3, A5	✓	✓	✓	✓	✓
At5g18460	H/I2, E/F9, G/H11	H2, F9, G11	✓	✓	✓	✓	✓
At1g69690	D1, G9	D1, G9	✓	✓	✓	✓	-
At5g06250	D3, I10, I11	D3, I10	✓	✓	✓	✓	-
At5g46590	F2, G2,	G2	✓	✓	✓	-	-
At1g24020	F3	F3	✓	✓	✓	-	-
At1g78170	D4	D4	✓	✓	✓	-	-
At1g62540	I5, A12	I5, A12	✓	✓	✓	-	-
At1g19830	D6, E6	E6	✓	✓	✓	-	-
At1g31310	H4, I9	I9	✓	✓	✓	-	-
At1g65370	E5, F5, J10, A11	E5, F5, J10	✓	✓	✓	-	-
At1g77200	J2, A3	J2, A3	✓	✓	✓	-	-
At1g02190	E4, F4	E4, F4	✓	✓	✓	-	-
At1g73620	J5, A6	J5	✓	✓	✓	-	-
At1g71050	B9	B9	✓	✓	✓	-	-
At5g19730	F8	F8	✓	✓	-	-	-
At2g16210	E1, F1	E1, F1	✓	-	-	-	-
At5g24580	B3	B3	✓	-	-	-	-
At1g52410	E3	E3	✓	-	-	-	-
At1g54020	C5, D5	C5	✓	-	-	-	-
At3g55110	G5, H5, J11	G5, H5	✓	-	-	-	-
At4g27590	B7	B7	✓	-	-	-	-
At5g57130	C7	C7	✓	-	-	-	-
At2g32280	H7	H7	✓	-	-	-	-
At1g62500	G8, H8	G8, H8	✓	-	-	-	-
At1g60060	A2, B2, H9	A2, B2, H9	✓	-	-	-	-
At3g28220	B10	B10	✓	-	-	-	-
At1g03710	I4, J4	I4	✓	-	-	-	-
At2g39330	F6, G6	F6, G6	✓	-	-	-	-
At3g12460	E7, F7	E7, F7	✓	-	-	-	-
At5g44620	G7	G7	✓	-	-	-	-
At5g51850	I7	I7	✓	-	-	-	-
At4g02670	J7	J7	✓	-	-	-	-
At3g13662	B8, C8	B8	✓	-	-	-	-
At5g49330	G3, H3, J9	G3, H3	✓	-	-	-	-
At1g24070	D7	D7	✓	-	-	-	-
At5g10420	C4, E11, F11	C4	✓	-	-	-	-
At4g25020	B6, C6	B6	✓	-	-	-	-
At1g03720	H6, I6	H6	✓	-	-	-	-
At1g51460	J6, A7	A7	✓	-	-	-	-
At2g16580	I8, J8	I8, J8	-[1]	✓	-	-	-
At2g01610	A13	A13	-[2]	-	-	-	-
At5g20740	A9	-[3]	-	✓	-	-	-
At4g16447	-[4]	-	-	✓	✓	✓	
At3g15680	-[4]	-	-	✓	-	-	-
Sum	44 (92)	43 (67)	41	21	17	6	3

[1] I8 did not germinate in SD, J8 currently tested
[2] A13 line just recently available
[3] No T-DNA detectable by PCR
[4] No T-DNA insertion lines available

Table 5.3: **Summary of screening progress for candidate genes 2.**

Gene	Lines ordered	Homozygous lines	SD	OE T_0	OE T_1	OE T_2	OE T_3
At5g17030	A8	-[1]	-	-	-	-	-
At1g28290	C9	-[2]	-	-	-	-	-
At2g28790	D9	-[2]	-	-	-	-	-
At1g26780[3]	J1	J1	✓	-	-	-	-
At1g13290[4]	A4	A4	-	-	-	-	-
At1g70560[5]	D8, E8	E8	-	-	-	-	-

[1] Seeds did not germinate, one plant recovered from a GM plate, progeny currently analysed
[2] No T-DNA detectable by PCR
[3] Identified as LATERAL ORGAN FUSION1 (LOF1) (Lee et al., 2009)
[4] Identified as DEFECTIVELY ORGANIZED TRIBUTARIES 5 (DOT5) (Petricka et al., 2008)
[5] Identified as TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS 1 (TAA1) (Stepanova et al., 2008)

5.5 Evaluation of DPAs

All T-DNA lines for which an abnormal phenotype could be observed were further analysed to confirm the stability of the phenotype and the connection between the annotated T-DNA insertion and the observed phenotype. A phenotype was considered stable when it was observed in at least two independent plantings of the homozygous seeds of a line. This was done to rule out instabilities in the conditions (e.g. herbivores) as a cause for the differences to wild type. For the three T-DNA lines where the only observed phenotype was early flowering in SD, no repetition has been done so far since the phenotypes were so mild and they will only be followed up if confirmed by another line. 50 % of lines from SALK collection contain more than one T-DNA insertion (http://signal.salk.edu/tdna_FAQs.html). Thus one T-DNA insertion line displaying an abnormal phenotype is not sufficient to confirm the connection between the gene and the phenotype. A phenotype was considered confirmed, if observed in two or more independent T-DNA insertion lines or in two or more independent overexpressor lines (Tab. 5.4). Observations of converse phenotypes in a T-DNA insertion line and an overexpressor line was a second option for confirmation. T-DNA insertions were also assessed for reduction of transcript as result of the insertion. The following section describes the evaluation and the status of analysis for each DPA except DPA6 (5.4.3).

Table 5.4: **Summary of observed phenotypes.**

Name	Gene	Type[1]	1^{st} screen LD	1^{st} screen SD	Overexpressor (OE)	Stable?[2]	Confirmed?[3]
DPA1	At1g75710	C2/H2	late flowering (C2)	late flowering (C2)	no phenotype	✓(C2)	x
DPA2	At5g18460	unknown	no HM[4] (I2/E9)	no phenotype (F9 tested)	no phenotype	✓(I2/E9)	pending
DPA3	At1g65370	MATH	short petals (A11)	no phenotype (J10 tested)	T1 no phenotype in LD	✓(A11)	x
DPA4	At5g06250	ABI3_VP1	no phenotype	stronger leaf serration, enlarged petals (D3, I10)	no leaf serrations, dwarf, fasciated stem, misshaped siliques	✓(D3, I10)	✓
DPA5	At1g69690	TCP	bushy, small (D1)	some plants small (D1)	no phenotype	✓(small)	✓(G9)
DPA7	At5g10420	Antiporter	no phenotype	slightly early flowering (C4)	n.d.[5]	n.d.	pending
DPA8	At1g78170	unknown	no phenotype	slightly early flowering (D4)	T1 no phenotype in LD	n.d.	pending
DPA9	At5g51590	unknown	no phenotype	no phenotype (I3 & J3)	OE small & early flowering (LD & SD)	✓(OE)	✓
DPA10	At1g02190	CER1	no phenotype	early flowering (E4)	n.d.	n.d.	pending
DPA11	At3g55110	ABC transporter	no phenotype	small (H5)	n.d.	✓(H5)	✓
DPA12	At1g62540	FMO GS-OX2	no phenotype	more cauline leaves, apical dominance reduced (I5)	T1 no phenotype in LD	✓	pending
DPA13	At1g03720	cysteine-type peptidase	no phenotype	late flowering (days) and small (H6)	n.d.	x	x
DPA14	At1g24070	CSLA10	no phenotype	1/3 of plants: dwarf, dark round leaves, late flowering	n.d.	✓	pending

[1] Function or family affiliation of proteins assigned by sequence similarity, for abbreviations see text
[2] Phenotype was reproducible in following sawings
[3] A phenotype is considered confirmed if two or more lines show it, or two or more overexpressor lines show it
[4] No homozygous plants found during genotyping, I2 and E9 are derived from the same T-DNA insertion
[5] Not done

5.5.1 Candidate genes excluded due to instability of phenotype or missing confirmation

The genes mentioned in this section were excluded from the further analysis because of an instable phenotype or a clear evidence that the phenotype is linked to another gene:

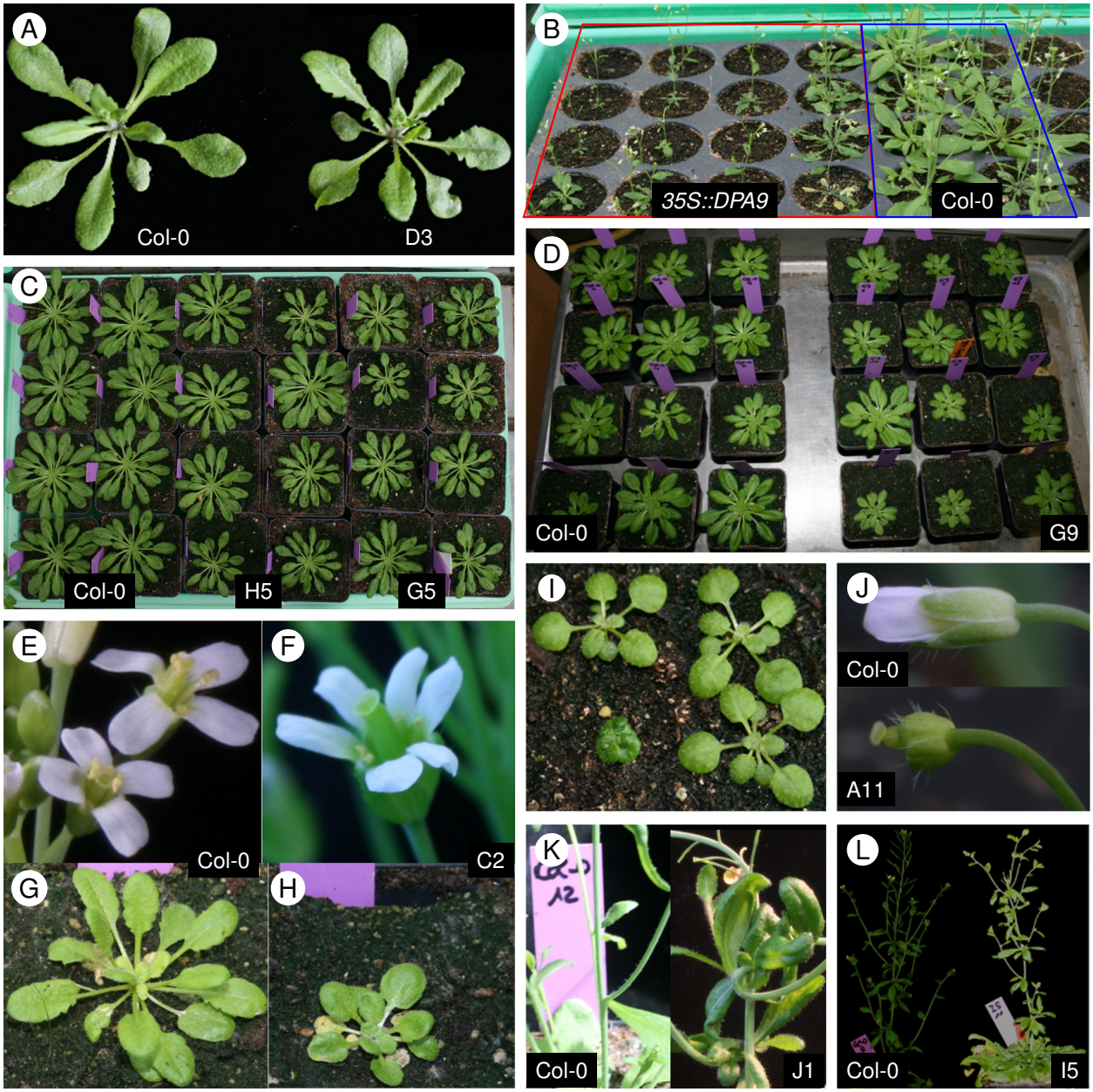


Figure 5.7: **Phenotypes observed in screen.** **A**: DPA4 - LD grown plants of Col-0 and line D3, leaves of D3 plants display enhanced serration. **B**: DPA9 - *35S::DPA9* plants compared to Col-0 plants in LD conditions, mutants are smaller and flower earlier compared to wild type (wt). **C**: DPA11 - SD grown plants for Col-0 and two insertion lines in DPA11 locus. Rosette diameters are smaller compared to wt in both lines. **D**: DPA5 - comparison of Col-0 plants to insertion line G9, some plants in line G9 are smaller than wt. **E-H**: DPA1 - flowers and vegetative rosettes of Col-0 and line C2 in LD, flowers display higher numbers of petals in C2, rosettes are smaller and leaves are round and dark green. **I**: DPA14 - segregation of line D8 in SD. About one-third of the plants are dwarfish with round, dark green leaves like the plant in the lower left corner, all other plants are indistinguishable from wt. **J**: DPA3 - flowers of LD grown plants from line A11 and Col-0. Petals in line A11 are smaller than sepals. **K**: DPA6 - line J1 curls side shoots when grown in SD conditions. Insertion of J1 was published as an allele of *lof1* (Lee et al., 2009). **L**: DPA12 - comparison of line I5 to Col-0 in SD conditions: in I5 plants the main shoot fails to elongate after initiation of the side shoots and remains shorter than those.

DPA1 - At1g75710

The first tested line for this gene (C2) displays late flowering in LD and SD, round, dark green leaves and abnormal numbers of floral organs (Fig. 5.7 E). Further analysis revealed that also trichome branching and epidermal cell shape is altered in this line. Trichomes show four branches more often than in wild type and also some cases of five branches as well as twin-trichomes were observed. Epidermal cells are rounder and smaller in C2 line. However, no other insertion line in the *DPA1* locus displayed this phenotype. Therefore, C2 plants were crossed back to Col-0 plants and the progeny was analysed and genotyped. This experiment clearly revealed that the phenotype is not caused by the C2 insertion and the gene was excluded from further analysis in this project. Since the phenotype is nevertheless interesting it will be further analysed in future. So far four other T-DNA insertion could be found in this line but none was responsible for the altered development.

DPA3 - At1g65370

One insertion line (A11) was noticed in the screen because of its small flowers (Fig 5.7 G). Sepals are to some extent misshapen in A11 plants, carpels and petals reduced in size. No other insertion in this locus displayed the same phenotype and expression analysis in all available lines revealed strong reduction of *DPA3* in line J10 but not in A11. Due to this anticorellation of expression pattern and phenotype, expression of the neighbouring gene *CLAVATA2* (*CLV2*) was analysed in A11. *CLV2* regulates meristem and floral organ development and might therefore be responsible for the alteration in floral organs of A11. Loss of CLV2 function results in enlarged meristems and increased floral organ numbers (mostly carpels and stamens) (Kayes and Clark, 1998). A11 is located in the promoter of *DPA3* and due to head-to-head arrangement of *DPA3* and *CLV2* A11 might also influence *CLV2* expression. Expression of *CLV2* is indeed altered in A11, a two-fold upregulation was observed. Therefore, one could assume that the observed defects in floral organs are caused by *CLV2* overexpression and *DPA3* was not considered for further analysis. A recent study reported no abnormal phenotype for plants overexpressing *CLV2* (Wang et al., 2010), suggesting that other factors might play a role in A11 plants, maybe another T-DNA insertion.

DPA13 - At1g03720

Line H6 displayed small rosette diameter and a delayed bolting date (leaf number was not altered) compared to wild type in the first SD screen. This phenotype was not observed in following plantings. This and the fact, that there are no other homozygous T-DNA line available (in a second available line, I6, no T-DNA insertion could be detected by PCR) led to the exclusion of *DPA13* from further analysis.

5.5.2 Candidate genes for which confirmation is pending

This section describes the observation for genes with not yet confirmed phenotypes. Since they are not confirmed yet, the causes and underlying genetic networks of these mutations have not been analysed in detail so far.

DPA2 - At5g18460

For one T-DNA insertion line (I2) in the *DPA2* locus no homozygous plants could be found during the genotyping progress and in following plantings, neither on soil nor on GM medium plates. In the mean time a putative homozygous version of line I2 (named I9 in this screen) was ordered from the SALK institute but also no homozygous plants were found in this line. The correct annotation of the insertion site was confirmed by sequencing of the LB/RP PCR-product from genotyping. Of four additional lines analysed in this locus three are available in homozygous form (H2, F9 and G11), one (H11) did not germinate on soil and GM medium. Expression analysis of F9 and G11 revealed that they are partial knock outs of *DPA2* and the complete transcript is still detectable in these lines (H2 analysis currently in progress). Therefore, it might be that I2/E9 and H11 are complete knock outs and are not viable in homozygous form. To test this hypothesis, the genomic region of *DPA2* will be cloned and transformed into I2 plants. This construct should complement the mutation and enable homozygous maintenance of the insertion, if the insertion is really lethal.

DPA7 - At5g10420

Line C4 was observed to be slightly early flowering in SD. Two other insertion lines in the *DPA7* locus (E11 and F11) are rather later flowering than wild type, which might be

caused by overexpression of *DPA7* either in C4 or in E11/F11. Expression will be evaluated in all three lines, but since the effect was very mild in all cases and not significant for the tested sample size, this gene will probably be excluded from further analysis.

DPA8 - At1g78170

Line D4 was also observed to be slightly early flowering in SD. Since there are no other T-DNA insertion lines available, an overexpressor was created. Several independent lines did not show any phenotype in LD in T_1 generation. Further generations will be analysed in LD and SD conditions.

DPA10 - At1g02190

Line E4, an insertion line in the promoter of *DPA10*, flowered significantly early in SD. Another promoter insertion (F4) did not show any abnormal phenotype. The full-length transcript of *DPA10* is still detectable in both lines, the amount of amplified transcript suggests that E4 produces rather more *DPA10* transcript than wild type. This result has to be confirmed by qRT-PCR. Furthermore the expression of the neighbouring gene (At1g02180, an unknown ferredoxin-related protein) will be tested. As a putative ECERIFERUM1 (CER1) protein one would expect DPA10 to be involved in wax biosynthesis, but no obvious defect in wax was observed. A transcriptional profiling study suggests a role of *DPA10* in floral organ development, since *DPA10* was found to be upregulated during *AGAMOUS* (*AG*) induced stamen and carpel initiation in *AP1* and *CAULIFLOWER* (*CAL*) double mutants, which fail to initiate floral organs (Gómez-Mena et al., 2005). However, no defects in floral organ development were observed in E4 and F4 plants.

DPA12 - At1g62540

An exon insertion line in *DPA12* displayed reduced apical dominance in SD, leading to a non elongated main inflorescence and a higher number of cauline leaves compared to wild type (Fig. 5.7 I). A promoter insertion line does not show any abnormal phenotype. Therefore, expression of *DPA12* will be tested in both lines. In parallel an overexpressor was created but did not display any phenotype in LD in T_1 generation. After the beginning of this screen, an enzymatic activity of DPA12 was reported, proving

that DPA12 has indeed FLAVIN-MONOOXYGENASE (FMO) function with GLUCOSINOLATE S-OXYGENASE (GS-OX) activity as predicted by sequence similarity before (Li et al., 2008b). The authors of this report also used the T-DNA insertion lines named I5 and A12 here and state that they are knock outs of *DPA12* expression. Furthermore an overexpressor was created by Li et al. (2008b) to test enzyme activity. Neither for the T-DNA lines nor for the overexpressor any developmental phenotype was reported, but plants were only analysed in LD conditions. If I5 and A12 should be complete knock outs of *DPA12* expression the phenotype observed for I5 is likely to be caused by another T-DNA insertion.

DPA14 - At1g24070

About one-third of the plants from line D8 are dwarfish with round, dark green leaves (Fig. 5.7 F) and extreme late flowering in SD. The plants did not flower after three months. To obtain seeds, they were then shifted to LD conditions where they immediately started to flower and produced a bushy shoot. The segregation ratio could be a hint to another transgene which might cause the phenotype or might be caused by the penetrance of the phenotype. More insertion lines will be analysed and progeny of the plants with both the phenotype and without will be planted again. DPA14 was annotated as CELLULOSE SYNTHASE LIKE A10 (CSLA10), suggesting a role in cellulose biosythesis. A hint that DPA8 might be indeed the gene affected in the dwarf D8 plants is the fact that these plants resemble the phenotype of cellulose synthase catalytic subunits (CESAs) double mutants (*cesa2/cesa6* double mutant (Desprez et al., 2007)).

5.5.3 Candidate genes with confirmed associated phenotype

DPA4 - At5g06250

Two independent T-DNA insertion lines in the *DPA4* locus display enhanced leaf serration (Fig. 5.7 A) and enlarged petals. This phenotype was first observed in SD but is also visible in LD, where it was overlooked in the first screen. *DPA4* encodes a putative B3 domain containing transcription factor and because of its confirmed phenotype, *DPA4* was analysed in detail. This analysis is described in Part II.

DPA5 - At1g69690

DPA5 is a putative TEOSINTE BRANCHED1/CYCLOIDEA/PCF (TCP) transcription factor and was added to the list of DPAs because of the bushy, small phenotype displayed by the T-DNA insertion line D1 in the first LD screen. Later on only the size phenotype was found to be stable, bushy plants were never observed again. Some plants of D1 and a second line (G9) display smaller rosette diameters when grown in SD conditions (Fig. 5.7 D). TCP transcription factors are already known to be involved in leaf shape and size regulation in *Antirrhinum* (Nath et al., 2003) and were shown to be the cause of *miR319a* associated leaf shape alterations in Arabidopsis (see 8.2.1). Currently, TCP transcription factors are also analysed for a role in leaf size regulation in Arabidopsis (Matser et al., 2010).

DPA9 - At5g51590

DPA9 was identified during the overexpressor screen. Several independent lines of the *35S::DPA9* construct display significantly smaller rosette diameter and early flowering in LD (Fig. 5.7 B) and SD conditions compared to wild type. Exact determination of flowering time revealed average differences in total leaf number of 2 leaves in LD and 10 leaves in SD conditions (Fig. 5.8 C and D). Differences in rosette diameter at time of bolting were 3.5 cm on average (Fig. 5.8 B), the lower diameter in *35S::DPA9* plants was partly caused by shorter petioles of rosette leaves. DPA9 was one of the first candidates to be included in the overexpressor screen due to its similarity to the gene product of the At4g25320 locus (62.9 % identity on amino acid level). None of the three available homozygous T-DNA insertion lines (I3, J3 and A5) displayed any growth abnormalities, although *DPA9* expression was found to be strongly reduced at least in J3 (Fig. 5.8 A). Together with the strong phenotypic alteration in the overexpression line this confirms the suspicion that DPA9 and the gene product of At4g25320 might act redundantly. To test this hypothesis, T-DNA insertion lines in the At4g25320 locus were ordered and crossed to J3 plants. Plants homozygous for both insertions will be analysed in future. The sequence of DPA9 indicates that DPA9 might act as a DNA binding protein, since it contains an AT hook (DNA-binding motifs with a preference for A/T rich regions) (Meijer et al., 1996). This suggest a role of DPA9 in the regulation of transcription which

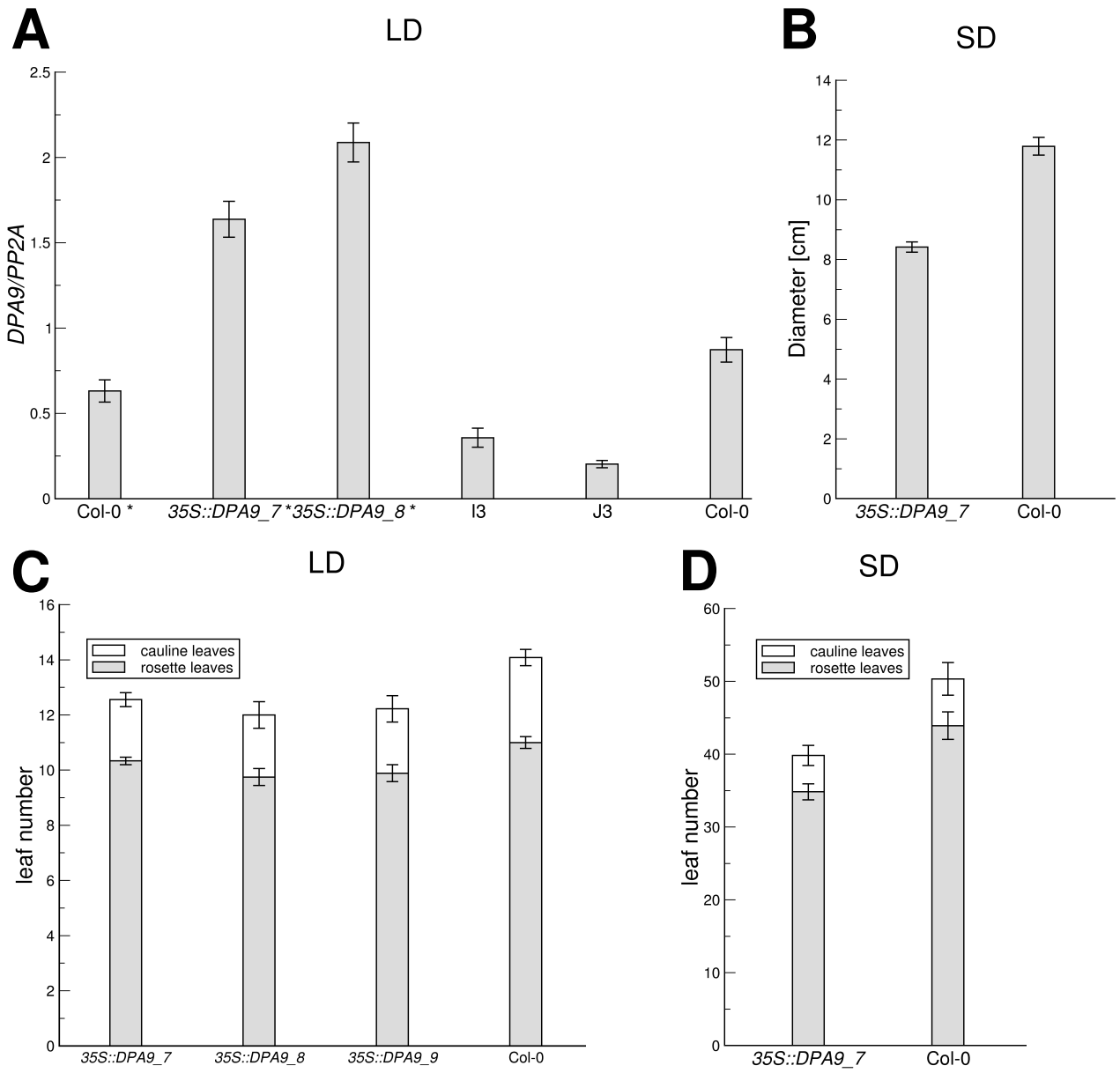


Figure 5.8: **Phenotype of *35S::DPA9* plants.** **A**: Expression of *DPA9* at ZT8 under LD conditions in two homozygous populations of the same line (*35S::DPA9_7*) and two T-DNA insertion lines located in the *DPA9* locus. Samples marked with asterisk belong to a different experiment and are only comparable to the respective control. Data is based on a single experiment. **B**: Rosette diameters of Col-0 and *35S::DPA9_7* plants at time of bolting in SD. **C**: Flowering time of three homozygous populations from line *35S::DPA9_7* in LD conditions. Data is based on a single experiment, similar tendencies were observed for several independent lines. **D**: Flowering time of one homozygous population from line *35S::DPA9_7* in SD conditions. Data is based on a single experiment. All error bars represent standard errors of the mean.

will be further analysed by qRT PCR of known regulators of flowering time in future experiments. *35S::DPA9* plants are still able to perceive day length, flowering occurs much later in SD than in LD, although earlier than wild type in both cases. This suggests a role of DPA9 in a rather photoperiod independent manner. One possible pathway where DPA9 might be involved in is the GA mediated flowering time pathway, which promotes flowering under SD conditions (Wilson et al., 1992), mainly through the activation of the floral meristem identity gene *LFY* (Blazquez et al., 1998). Therefore, *LFY* and genes in the GA mediated flowering time pathway will be the first candidates to test expression.

DPA11 - At3g55110

Line H5 (exon insertion) displayed significantly smaller rosette diameter than wild type at time of bolting when grown in SD conditions. This phenotype was first observed in the SD screen (growth cabinet) and was confirmed in SD conditions in a greenhouse. Here the plants also flowered significantly earlier than wild type (about 7 leaves less than wild type). To confirm the phenotype, a second exon insertion line, G5, was planted side by side with H5 and Col-0 in SD conditions. This preliminary experiment was only done with 8 plants each but revealed a tendency of both H5 and G5 to smaller rosette diameters and earlier flowering. The differences in rosette diameter are more obvious before bolting (Fig. 5.7, C). This preliminary result seems to confirm the linkage of the phenotype to the *DPA11* locus. Because of sequence similarity, DPA4 is supposed to be an ATP-binding cassette (ABC) transporter family protein. This rises the question whether the small size of the plants can be considered as a developmental effect or is a result of starvation to certain substances that are no longer transported in the mutant. Further investigations will address this question.

Functional analysis of PcG targets genes via Gene Ontology terms had already indicated an overrepresentation of development related genes, especially in sub sets of target genes expressed in the same tissue (Zhang et al., 2007a). Although enriched in well characterised genes that play a role in development, these sub sets contain a high number of still uncharacterised genes. Therefore, this study aimed to test whether still uncharacterised genes of a sub set of PcG target genes expressed in the apex and floral organ also play a role in development. The aim is to identify novel genes that contribute to shoot development in Arabidopsis.

Although shoot development has been widely studied and many genes involved in developmental processes are already characterised, there still seem to be several missing links between regulatory components, *e.g.* in the regulation of flowering time and organ initiation in the meristem (Turck et al., 2008; Barton, 2010). Some missing regulators might have been missed by classical forward genetics screens due to existing redundancies, whereas forward genetics where single loss-of-function mutants are likely to be discovered are saturated to some extent.

The candidate based reverse genetics approach in this study allows the generation of selected overexpressors for putative redundant genes. Furthermore mutants for the relative small set of genes can be screened in different conditions (*e.g.* LD and SD) and different traits, which might be quite laborious for the usual number of ten thousands of mutagenised plants in a forward genetics screen. Therefore, in contrast to classical forward genetics screens for missing regulators in a certain process of shoot development, the reverse genetics approach in this study was designed to screen for abnormal phenotypes that affect several stages of shoot development. Traits scored during the screen were flowering time, leaf shape, phyllotaxis, size and inflorescence architecture. An advantage over the forward genetics method is that once an abnormal phenotype is observed, the associated locus is already known and the gene can be analysed in detail without prior mapping.

6.1 Choice of PcG target gene set

One critical decision for the complete work was the choice of the PcG target gene set. All so far available data sets were obtained from seedlings. It could be argued that many target genes are not detected, if for example a gene is only targeted in later stages of plant development. Nevertheless, several characteristics of the H3K27me3 targets argue against this concern: PcG targets are in general expressed at a lower level compared to an average gene and their expression is very tissue specific (Zhang et al., 2007a). These observation established the current model that PcG targets are generally repressed throughout development and that their repression is only released in very specific tissues and stages. Therefore, it is much more likely to find a target in a repressed state marked by H3K27me3 than finding a target locus in a not repressed state, especially if a mixture of tissues is used as in the case of seedlings.

The published data set (H3K27me3_2007, 4979 genes) from Zhang et al. (2007a) was generated using Affymetrix®tiling arrays. Our preliminary data at this time point suggested that the signal intensities in our NimbleGen tiling arrays were slightly higher and displayed a better signal to noise ratio. Therefore, we decided to use our data set for this study, although it was generated for LHP1 in L*er* (LHP1_2007, 5057 genes). 72% of the genes in the LHP1 data set were present in both sets, but this rises the question whether the remaining 28% were true positives. This can be partially answered by comparison of both sets to the recent H3K27me3 target list for Col-0 generated in our group (H3K27me3_2010).

H3K27me3_2010 contains 7245 gene and would be used if the study was started now. 89.7% of LHP1_2007 genes were present in the new list and 86.7% of H3K27me3_2007. This in the end ascertains that choosing the LHP1 data set did not result in a greater deviance from H3K27me3_2010 than using the published data set. The overlap seems to be even slightly better for LHP1_2007, but this fact was not predictable in 2007. It should be anyway noted that the overlap between the lists of about 90% is quite close to the overlap we usually observe between replicates.

To evaluate the outcome of this study, it is also interesting to know the overlap of the recent target list with the identified apex cluster. Three genes of the apex cluster were not present in the new list, one of them was already known and therefore not considered

as a candidate. The two remaining genes are intermediate cases where it is not possible to determine if they are targets based on visual observation in a browser. In both cases enrichment is mainly abundant in the promoter, a fact that was considered in the LHP1_2007 list (genes were defined as coding region plus 1000 bp upstream and 200 bp downstream) but not in the recent analysis.

Another question resulting from the new H3K27me3 target list is, whether it would make sense to repeat the analysis conducted in this study with the larger H3K27me3_2010 list. H3K27me3_2010 was generated using a different analysis method and thus a differently defined threshold for positive genes resulting in the higher number of genes. The relatively low overlap between LHP1_2007 and H3K27me3_2007 (72 %) and the high overlap of both to H3K27m3_2010 indicates a number of false negative genes in both old lists. However, this also suggests that the LHP1_2007 list was very restrictive and probably contains less false positive genes than the larger H3K27me3_2010 list. Therefore, it would make sense to repeat the analysis for the additional genes but the resulting candidate genes should be carefully analysed in a browser for the H3K27me3 mark before analysing T-DNA insertion lines since a larger list will always contain more false positive genes than a more restrictive one.

6.2 Choice of expression set

The "Developmental series" data set was used in two versions which varied only in the normalisation method. Using both data sets resulted in a very restrictive set of only 105 apex expressed genes, whereas using only today's standard method gcRMA alone would have yielded twice as many genes. As 100 candidates would have been a large number to screen, a subdivision would have been necessary anyway to decide which genes to start with. Since the middle of the year 2008 a developmental expression set generated with a whole genome tiling array called *Arabidopsis thaliana* Tiling Array Express (At-TAX) is available (Laubinger et al., 2008). Usage of this data set in future would result in a higher number of genes included in the analysis, for example 6794 genes from the recent Col-0 list are included in that data set in contrast to the 4872 included in the ATH1 derived data set. However, the new developmental expression set consist only of 11 samples (in contrast to 79 in "Developmental series") and only two of them are apex samples (in

contrast to 4 wild type and 7 mutant samples in "Developmental series", see section 4.2). Therefore, the "Developmental series" data set provides a more robust tool for clustering and identification of tissue specific expression patterns and should be also used in future for genes present on the ATH1 array. At-TAX data can be used for the remaining genes to enable their inclusion in further analysis.

6.3 Candidate gene selection procedure

6.3.1 Selection and functional analysis of an apex expressed PcG cluster

In this study, a cluster containing genes expressed in the shoot apex and floral organs was obtained and a high enrichment of developmental GO terms could be observed in this cluster. It could be demonstrated that this enrichment is specific to apex expressed H3K27me3 targets and could not by obtained by only selecting for apex expressed genes or for PcG targets.

The functional enrichment analysis performed in this study employed all GO terms, including those which were interfered from structural similarity. Therefore, also some of the not yet characterised genes contributed to the analysis because of their similarity to other, already known genes. These terms are of course less reliable, but it should be noted that only very general GO terms are assigned that way. In the particular case of this study only the terms concerning transcription were affected since transcription factor activity is quite often interfered from sequence similarity. The GO terms concerning development were all annotated based on experimental data.

6.3.2 Selection of candidates from apex cluster

To select so far uncharacterised genes among the apex cluster mainly the information provided by TAIR was used. This consortium provides a summary of the known aspects for each gene, including GO terms and publications. Still, each gene had to be analysed separately. It would have been of great benefit to automate this procedure, but for this a prerequisite would have been a complete and up-to-date database of all functions as-

signed to genes. The Gene Ontology is supposed to be such a database, but delays in annotation can not be prevented, since all the annotations have to be approved by the curating organisation. For this reason GO terms were only considered for a pre analysis and the results were confirmed by searching for available publications for each gene. A challenge for the future would be to create an at least partially automated selection procedure for genes with no known function in development, maybe by combining information from some of the existing web tools.

The selected gene set could be further subdivided in a set of genes almost only expressed in the shoot apex and floral organs (54 genes) and in a set with expression in all tissues but higher expression in the shoot apex and floral organs (51 genes). This distribution reflects the tissue specificity of PcG targets and it might be possible that the PcG specific characteristics tissue specificity and involvement in development are linked and that tissue specific genes are more often involved in development. Therefore, it might have been reasonable to only analyse the more tissue specific sub set.

With the results of the screen it is possible to decide whether the decision to include both sets made sense. After detailed review of the characterisation status, the same number of genes (26) from each list were included in the screen, indicating that many functions among the alleged well characterised set were assigned by computational analysis. Therefore, GO terms with purely *in silico* based evidence codes should have been excluded from this part of the functional analysis.

Likewise, the distribution of abnormal phenotypes observed per sub set is equal (7 phenotypes observed in each list). Of the genes with confirmed phenotypes in this study, only *DPA4* was found in the more exclusively apex expressed list. However, all three genes described by other research groups during this study are in this set. Taken together this analysis suggests that the exclusively apex specific group contained slightly more candidates for a role in development, but developmental function were not exclusively found in the more tissue specific set. Since the other group contained at least two candidates with a connection to development (*DPA5* and *DPA9*), it was correct not to exclude them from the screen.

6.4 Screening procedure

To identify relevant developmental regulators among the candidate genes, this study was first focused on the analysis of T-DNA insertion lines. Overexpressors were only generated in a second screening approach. Mutants were first screened in LD and afterwards in SD conditions. In the following section this procedure is discussed and potential drawbacks are elucidated.

6.4.1 Loss-of-function mutants vs. overexpressors

Abnormal phenotypes of loss-of-function mutants allow a more direct conclusion about the affected gene than overexpressors, which can also cause artefacts with no biological relevance. For example a transcription factor may bind to non target genes because of its high abundance or activate targets in ectopic tissue. However, overexpressors can reveal functions of genes despite existing redundancies. Ideally, overexpression confirms phenotypes observed in a loss-of-function mutant by causing an adverse effect. The strong developmental defects in PcG mutants are also mainly caused by upregulation of their target genes (Katz et al., 2004; Schubert et al., 2006), while ectopic expression of for example LHP1 does not result in developmental defects (Turck et al., 2007). Therefore, in the particular case of PcG target genes, which are thought to be generally repressed and only activated in a few tissues and stages, overexpressors might be superior to loss-of-function for the detection an abnormal developmental function. One would expect a loss-of-function mutant to only show defects in the stage and tissue, where the gene is usually expressed, which might be easily overlooked or not important in the particular screening conditions. In conclusion both loss-of-function mutants and overexpressors are likely to be beneficial and the screen had been probably faster and more successful if overexpressors would have been created for all candidates in parallel with the T-DNA insertion line screen.

6.4.2 Drawbacks of T-DNA insertion lines

As about 50 % of T-DNA insertion lines contain more than one insertion, phenotypic alterations observed in T-DNA insertion lines always have to be confirmed (see 5.5). The

analysis of *DPA1* made clear that up to five T-DNA insertions can be present in one line. Proving which T-DNA insertion line causes the phenotype can be a long procedure including outcrossing and functional complementation. Hence, the fastest and safest way to confirm the link between an annotated insertion and an observed phenotype is the analysis of a second T-DNA insertion line in the same locus.

Another drawback of T-DNA insertion lines is that they are not necessarily loss-of-function mutants for the gene carrying the insertion. In particular lines with promoter insertions lines but also some intron and exon lines, still produce the full-length transcript (*e.g.* lines F9 and G11 in *DPA2*). Furthermore, promoter insertion lines can cause overexpression of the gene instead of expression reduction. Likewise, insertions can influence the transcription of neighbouring genes, such as observed in line A11 for *CLV2*.

Nevertheless, being aware of these drawbacks, T-DNA insertion lines provide a fast tool to screen many genes in a relative short time with no delay for the construction of transgenic plants.

6.4.3 Screening conditions

It is important to know whether the screening in SD conditions was advantageous, to decide in which conditions the remaining overexpressor lines should be screened.

Only 4 of the 14 phenotypic alterations observed in this study where initially detected in LD conditions. This can be explained by development reasons (slower development in SD, later flowering etc.) but also by the fact, that screens in SD conditions are probably not as saturated as LD screens. But there is also a subjective aspect to concern: subtle abnormal phenotypes were more likely to be discovered in SD conditions because the plants were more intensely monitored over a longer time. Subtle abnormal phenotypes missed in LD conditions were DPA4 and DPA6 (LOF1).

In conclusion, 7 of the 14 phenotypic alterations were only found in SD. All phenotypes observed in LD could also be observed in SD (except for line I2/E9 and line A11 which were not tested). Therefore, SD conditions should be used to continue the screen in the overexpression lines.

6.5 Evaluation of DPAs

Among the 52 candidates, four genes were confirmed to cause phenotypic alterations in mutants (DPA4, DPA5, DPA9 and DPA11), three genes have been excluded from further analysis (DPA1, DPA2 and DPA13). For six (DPA3, DPA7, DPA8, DPA10, DPA12 and DPA14) the confirmation is still in progress and one was published as *LOF1* by another group (DPA6).

Of the four confirmed phenotypes, three are clearly developmental phenotypes, whereas DPA11 could also be just involved in metabolism. DPA4 and DPA9 are currently analysed in detail (see Section 5.5 and Part II). These examples and the three recently published development related genes show that it is indeed possible to identify development involved genes among the apex cluster selected in this study.

6.6 Effectiveness of the screen

Genes with a developmental function may have been identified in a randomly selected list of genes. Therefore, the effectiveness of the screening procedure carried out here depends on the identification of more genes than one would expect to find in a list of 52 randomly selected unknown genes. For a rough estimate of the number of development involved genes expected in a randomly selected set, without actually performing the experiment, GO terms can be used: in 2007 when the screen started, 5.8 % of all genes had the GO term "developmental process" assigned. At the same time, 40.8 % of all genes had the term "unknown biological process" assigned. Assuming that among those unknown genes the same percentage would be involved in development, 14.2 % of all genes should be involved in developmental processes. Therefore, also 14.2 % of 52 randomly selected genes should be involved in development, which are 7.4 theoretically expected genes. The estimation does not take the fact into account that developmental trait have been intensively studied and screened for and these screens are saturated. For this reasons the actual expected number of development involved genes is likely to be lower.

Taking the three published genes with developmental function and the three confirmed phenotypes into account, there were for sure 6 genes with development related phenotypes among the 52 screened in this study. This is close to the expected value for a

random list, but two more genes would make it already more than expected for a random list. Since confirmations for 6 more genes with developmental phenotype are pending, 13 more overexpressors are to be created and 47 overexpressors were not yet screened in SD, it is very likely that the expected number for a random gene list will be exceeded.

In conclusion it is not yet completely clear that the screen was effective, but there are indications that this might be the case.

With the possible exception of early lethal genes, classical forward genetics screens for macroscopically visible developmental functions in Arabidopsis are highly saturated. Forward genetics screens imply mapping of the affected gene after the identification of phenotypic alteration. After the sequencing of the Arabidopsis genome, many laboratories had engaged in reverse genetics screens to discover the function of important regulators in plants or of gene families. These efforts were based on sequence similarity and often futile because of expected genetic redundancy implied by the similarity. Therefore, recently performed gene function discovery efforts often relied on overexpressors, accepting the obvious drawbacks of the procedure (6.4.1). In the reverse genetics screen performed in this study a small, restrictive set of genes with high probability to be involved in development was selected and screened for a role in Arabidopsis shoot development. The selection was not based on sequence homology but on expression pattern and labelling by H3K27me3. In several cases, the observed abnormalities in shoot development could be directly associated with the affected genes. This allows direct characterisation of these genes in future. Beside DPA4, which is currently characterised in detail (Part II), three further genes with confirmed associated phenotype will be analysed in the future to understand their function in development.

For DPAs with pending confirmation for the association of the T-DNA insertion to the phenotype further T-DNA insertion lines and overexpressors will be analysed.

To complete the screen, overexpressors will be created for the remaining candidate genes and homozygous lines will be obtained for overexpressors. Although overexpressors might be a useful tool to discover functions among the generally low expressed PcG targets (6.4.1), it might be still informative to create double mutants of putative redundant genes. To address this question we will collaborate with the GABI-KAT duplo project, which will generate suggested double mutants with priority (www.gabi-kat.de/duplo.html).

Although the main contributors to the developmental defects in PcG mutants are already known, it would be interesting to find out whether the gene with developmental function observed during this study contribute to the developmental defects in PcG mu-

tants. To address this question, double mutants of confirmed DPAs and PcG mutants will be created. If the DPAs contribute to the developmental defects in the PcG mutants these defects should be partially reduced in the double mutants.

The fact that also other groups of tissue specifically expressed genes exist among the H3K27me3 targets suggest that they could also be enriched for development associated genes. This could indeed be proven for a set of seed specific expressed genes, which display a strong overrepresentation for embryo development associated GO terms (Engelhorn and Turck, 2010). Putative loss-of-function alleles are currently screened for developmental abnormalities in the group of Professor Justin Goodrich at the University of Edinburgh.

The analytic pipeline described in this study (transcriptional clustering of PcG target genes) could also be expanded to other species. Since PcG mediated repression is highly conserved it is very likely to control developmental processes in a similar way in other higher plants. Thus it should be possible to detected candidates for a function in development of other species. The only prerequisites for this study are genome-wide PcG target gene data and expression data for a few different tissues. In the era of next generation sequencing this information can be obtained for any species with a sequenced genome by ChIP followed by sequencing (ChIP-Seq) and transcriptome profiling by RNA-Seq. Since no plant species has been screened for gene function to a similar extent as Arabidopsis, the screen will probably yield more candidates in other organisms, especially in recently sequenced ones.

Part II

Analysis of DPA4

8.1 Importance of leaf shape regulation

The most important role of leaves is to ensure the photosynthetic supply of nutrients for the whole plant. In addition to that, leaves are responsible for most of the gas exchange of plants, the distribution of nutrients and water transport (Tsukaya, 2006). To realise this function in an economic way, the plant needs to reach maximum light supply with a minimal leaf surface. This can be regulated by leaf positioning, leaf size and leaf shape. Auxin regulates leaf positioning according to phyllotaxis, but also leaf size and leaf margin shape. In accordance to that, auxin maxima direct leaf initiation in the SAM and leaf and margin outgrowth during leaf development (Bainbridge et al., 2008; Scarpella et al., 2010) (Fig. 1.2).

8.2 Leaf margins in Arabidopsis

An important trait of leaf shape variation is the outline of the leaf margin. The leaf margin is characterised by highly elongated epidermis cells (Kawamura et al., 2010) and can be either smooth, serrated (containing small tooth-like structures) or lobed (large outgrowths from leaf margin). The degree of serrations differs between accessions and is plastic according to environment and developmentally regulated: juvenile leaves are smooth and serrateness is stronger in later produced leaves. Morphologically serrated leaf margins can be divided into the tooth regions and the region between teeth, which is termed the leaf sinus. Increased serration can either be caused by an increased number of teeth or an increased sinus depth (Nikovics et al., 2006; Tsukaya, 2006). Serrations are often associated with hydathodes, secretory structures at the endpoint of vessels that are predicted to facilitate excretion of toxic or not needed organic substances and salts (Pilot et al., 2004). Hydathodes also facilitate the invasion of pathogen into the vascular system of the plant and a high number of hydathodes might therefore be a potential threat for the plant (Hugouvieux et al., 1998). Hydathodes are usually located at the tip of teeth, but in some cases they occur independent (*e.g.* the apical hydathode at the leaf tip) (Candela

et al., 1999).

8.2.1 Serration formation in Arabidopsis

Initiation and formation of leaf serration is regulated by similar mechanisms as leaf initiation in the SAM (Fig. 1.2) (Scarpella et al., 2010). As in the case of leaf formation, the auxin efflux carrier PIN1 is required to establish auxin maxima that direct the outgrowth of the serrations (Scarpella et al., 2006). In accordance to that, loss of PIN1 function results in smooth leaf margins (Hay and Tsiantis, 2006). Recent studies suggest a role of *CUC2* in formation of auxin maxima, since auxin distribution was found to be no longer discrete but equal at the smooth leaf margin of *cuc2* mutants (Kawamura et al., 2010). *CUC2* in leaves is transcriptionally downregulated by the micro RNA *miR164a* and in accordance to that *miR164a* mutants display enhanced leaf serrations whereas overexpression of *miR164a* results in smooth leaf margins (Nikovics et al., 2006).

The role of *CUC2* in leaf development was described later than its tissue separation function and is unique to *CUC2*, whereas the other *CUC2* functions like organ boundary specification and meristem formation are redundant between the three *CUC* genes (Section 1.3.2). In accordance to that, although *CUC1* and four other *NAC* gene family members are regulated by *miR164a*, *CUC2* was found to be upregulated in *mir164a* leaves compared to wild type, while *CUC1* expression was not detectable (Nikovics et al., 2006). NAC1 and two other factors where slightly upregulated in the mutant. Expression of a *miR164a* resistant version of *CUC2* phenocopied *miR164a* mutant plants, indicating that *CUC2* is the main player in *miR164a* regulated leaf margin outgrowth. Nikovics et al. (2006) present as a possible explanation that a suggested gene duplication gave rise to *CUC1* and *CUC2* in Arabidopsis (Zimmermann and Werr, 2005), and that, after the duplication, *CUC2* might have developed its specific role in leaf development.

CUC2 and partially also *miR164a* are expressed at the leaf sinus, suggesting that they might be involved in preventing outgrowth at this position (Nikovics et al., 2006). Nevertheless, this was not found to be true, *CUC2* rather promotes outgrowth of teeth through promotion of cell proliferation, possibly through upregulation of the *KNOX* transcription factor *STM* (Kawamura et al., 2010).

Beside their role as meristem identity factors (Section 1.3.2), *KNOX* genes were al-

ready previously shown to generate lobed leaves when ectopically expressed in leaves (Cho et al., 2007). In wild type leaves, expression of *KNOX* genes in leaves is prevented by *AS1* and *AS2* as during leaf initiation (Hay and Tsiantis, 2006). Upstream of this regulatory network, *miR164* was suggested to be transcriptionally upregulated by the class II TEOSINTE BRANCHED1/ CYCLOIDEA/PCF (TCP) transcription factors, while *CUC* gene transcription is indirectly downregulated by class II TCPs in both a *miR164* dependent and an independent way (Koyama et al., 2007; Barkoulas et al., 2007). Abundance of *TCP2-4*, *TCP10* and *TCP24* mRNA is negatively regulated by *JAW-D* (*miR319a*) directed cleavage. Overexpression of *miR319a* causes serrate, crinkled leaves and defects in cotyledon development (Palatnik et al., 2003). Beside this, TCP3 has been shown to be involved in shoot meristem formations and overexpression of *TCP3* leads to plants with fused cotyledons, resembling *cuc1/2* mutant plants. This observation is consistent with both *CUC1* and *CUC2* being regulated by TCPs (Koyama et al., 2007).

Another factor regulating leaf margin formation is the zinc-finger protein SERRATE (SE), which was shown to ensure proper processing of pri-miRNAs (Yang et al., 2006). SE regulates expression of the HD-Zip III (class III homoedomain leucine zipper) genes *PHABULOSA* (*PHB*) and *PHAVOLUTA* (*PHV*) in the SAM and the adaxial part of leaves by taking part in the processing of *miR156* and *miR166*. *miR156* and *miR156* are required for transcriptional downregulation of *PHB* and in accordance to this, loss of SE function results in a reduction of mature *miR156* and *miR166* and thus an elevated HD-Zip III expression and expansion of the *PHB* expression domain (Grigg et al., 2005). The phenotype of the weak allele (*se*-1) includes increased leaf serration, defects in phyllotaxis and the transition from juvenile to adult phase. Stronger alleles also display altered adaxial/abaxial cell fate (Prigge and Wagner, 2001; Grigg et al., 2005). Adaxial/abaxial cell fate is controlled by HD-Zip III genes, gain-of-function of these genes leads to adaxialised leaves but not to serration of the leaf margin (Cho et al., 2007).

Since weaker *se* alleles also display phenotypes associated with *KNOX* gene misexpression, such as ectopic meristem and leaf stipule formation, SE was thought to regulate their expression as well. However, no misregulation of *KNOX* genes was observed in *se* mutants, which led to the suggestion that SE rather regulates the ability of shoot tissue to respond to *KNOX* genes (Grigg et al., 2005).

Leaf shape can be also determined by the rate of cell division in the leaf. For exam-

ple, overexpression of Kip-related proteins (KRP1/2), which reduce cell proliferation in leaves, results in enhanced leaf serrations and serrations of petals due to reduced number of leaf cells and concurrent enlargement of cells. KRP1/2 are cyclin-dependent kinase inhibitors (CKIs), which bind to cyclin-dependent kinases (CDKs) and inhibit their enzymatic function, thus leading to a slower cell cycling activity (Wang et al., 2000; Veylder et al., 2001; Nakai et al., 2006).

8.3 Regulation of petal size

Petals (and other floral organs) are believed to be transformed leaves and to be therefore regulated by similar mechanisms during their development (Honma and Goto, 2001; Tsukaya, 2006). It has been shown, that loss-of-function mutations in *miR319a* result in small, narrow petals (Nag et al., 2009), suggesting a regulation of petal size via TCPs analogous to leaf margin development.

The size of an organ can be either varied by varying the cell proliferation rate and thus the number of cells or by varying the cell expansion rate and thus changing the size of the cells. Positive regulators of cell division in petals are, among others, the lateral organ size controlling factors JAGGED (JAG) and ANT. The C_2H_2 type zinc finger transcription factor JAG was shown to restrict expression of the boundary controlling genes *CUC1* and *CUC2*, although the mechanism of JAG action is not clarified yet (Xu et al., 2008). *ANT* expression is in turn regulated by auxin (Section 1.3.2). An example for a negative regulator of cell division rate is the E3 ubiquitin ligase BIG BROTHER (BB), which was proposed to limit the duration of cell division period by degradation of growth activators (Disch et al., 2006). *DA1* (gene was named *DA1* because DA means“large” in Chinese) was also suggested to be involved in ubiquitin mediated growth inhibition due to its two ubiquitin interaction motifs (UIM), which are usually found in ubiquitin receptors (Hicke et al., 2005; Li et al., 2008c). Loss of function mutants in these genes also affect the size of other plant organs as sepals and stems (*bb*) and seeds and leaves (*da1*). Cell expansion rate and cell size are impaired in mutants of the basic helix-loop-helix transcription factor encoding gene *BIGPETAL* (*bpe*), which also display enlarged petals. *BPE* encodes two proteins (BPEp and BPEub), BPEp is produced by an alternative splicing event (intron retention). *BPEp* is expressed in petals and is responsible for the size variations in these

organs (Szécsi et al., 2006). Recent studies suggest that *BPE* acts downstream of the plant hormone jasmonate (Brioudes et al., 2009).

8.4 B3 domain transcription factors

The B3 domain is a conserved DNA binding domain that is only found in higher plants (Kagaya et al., 1999; Riechmann et al., 2000). Members of the B3 superfamily are involved in plant development, including hormone signalling and flowering time. A studied example is VERNALIZATION 1 (VRN1), which is involved in maintenance of vernalization-induced *FLC* repression after return to ambient temperatures. This function is dependent on the PRC2 protein VRN2 and the putative PRC1 member LHP1 (Gendall et al., 2001; Sung et al., 2006). Therefore, VRN1 might be involved in PcG protein mediated gene repression, but no direct evidence was found for this hypothesis so far (Levy et al., 2002; Farrona et al., 2008). Further well characterised B3 domain proteins are LEAFY COTELYDON 2 (LEC2), FUSCA3 (FUS3) and ABSCISIC ACID-INSENSITIVE 3 (ABI3), which are involved in embryo development and the ARF proteins that mediate auxin signalling (Suzuki et al., 2007).

In Arabidopsis, the superfamily consists of 118 proteins and it can be divided in five major classes: ABI3/VP1 (VIVIPAROUS1) family proteins, HSI (High-level expression of sugar-inducible gene) proteins, RAV proteins, members from the ARF proteins and REM (Reproductive Meristem) family proteins. All B3 proteins share a conserved B3 domain, which was first discovered in the Protein VP1 from *Zea mays* and in ABI3 in Arabidopsis. In addition to that, each family except the ABI3 family contains also additional domains, like an AP2 domain in the case of the RAV proteins and the AUX/IAA and Aux response factor domains in the case of the ARFs (Swaminathan et al., 2008; Romanel et al., 2009). For the ABI3/VP1, RAV and ARF family the hexameric DNA sequences, which are bound by the B3 domain are known: ABI3/VP1 proteins recognise a CATGCA motif found in many seed specific promoters (Mönke et al., 2004), RAV proteins recognise the sequence CACCTG via the B3 domain and the motif CAACA by the AP2 domain (Kagaya et al., 1999). ARF family members bind to TGTCTC auxin-response elements (AuxREs) in the promoters of auxin response genes (Ulmasov et al., 1999).

8.4.1 RAV transcription factors

The RAV family of transcription factors contains 13 member of which 6 carry an AP2 domain in addition to the B3 domain and 7 only contain a B3 domain. Already functionally characterised members of the RAV family are RAV1, TEM1, RAV2/TEM2, NGATHA 1 (NGA1), NGA2, NGA3 and NGA4. They are involved in regulation of growth and flowering time (RAV1, TEM1/2) and leaf shape and gynoecium development (NGA1-4) (Hu et al., 2004; Castillejo and Pelaz, 2008; Trigueros et al., 2009).

8.4.2 A transcriptional repressiv motif in B3 proteins

Among the B3 transcription factors a motif was identified that confers repressive activity. In B3 proteins the motif consists of a completely conserved core sequence (RLFGV) and some variable flanking amino acids (L/V RLFGV N/D M/L/V) and was found in 17 members of the superfamily, 11 of them are RAV proteins. The motif is also found in other transcription factors but there the first arginine in the conserved core can be replaced by lysine. Fusion of this motif to a transcriptional activator can turn the respective protein into a repressor (Ikeda and Ohme-Takagi, 2009).

9.1 Determination of sequence identity

Numbers of identical nucleotides or amino acids in two sequences were determined using European Molecular Biology Open Software Suite (EMBOSS) Pairwise Alignment Algorithms at the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) website (www.ebi.ac.uk/Tools). To find hexameric motifs in promoters the "Motif Analysis" tool from TAIR was used (www.arabidopsis.org). It compares the frequencies of hexameric motifs in the upstream regions of a submitted gene list to the frequencies of the motifs in the whole genome. The length of the upstream region considered for the analysis can be 500 bp, 1000 bp or 3000 bp. The program returns a *p*-value for each possible hexameric motif representing the probability of the motif to be overrepresented in the submitted gene set compared to the whole genome assuming a binomial distribution.

9.2 Plant material

PcG loss-of-function mutants *clf-28*, *swn7* and *clf-28/swn-7* were kindly provided by Prof. Justin Goodrich (University of Edinburgh).

9.3 Growth conditions

PcG mutants were grown on germination medium (GM) (0.5 strength Murashige and Skoog medium supplemented with 1 % sucrose) in growth cabinet (Percival) in controlled environment at 20°C in LD conditions of 16 h light and 8 h darkness. Light was provided by fluorescent tubes. Seeds were stratified at 4°C to synchronise germination.

9.4 Scanning electron microscope (SEM) imaging

Fresh plant material was frozen in liquid nitrogen and remaining surface water was sublimed. Afterwards, frozen samples were spattered with platinum and transferred to the microscope under constant vacuum (Zeiss SUPRA 40VP scanning electron microscope including cryopreparation and transfer system (EMITECH K1250X)).

9.5 *In situ* hybridisation

In situ hybridisation was performed as described in Bradley et al. (1993) with modifications described in Jang et al. (2009).

9.5.1 Harvesting and fixation of samples

"Apex" and "Inflorescence" samples were collected as described in 4.11 and directly fixed with 4 % (w/v) paraformaldehyde in Phosphate buffered saline (PBS, 0.15 M NaCl, 7 mM Na_2HPO_4, 3 mM NaH_2PO4 pH 7.0) containing 0.1 % Tween-20 and 0.1 % Triton X-100. To allow the formaldehyde to penetrate the cell walls, vacuum was applied twice for 10 min. Samples were left in fresh fixative over-night at 4°C on ice. Dehydration was performed in ethanol solutions in the following order:

30 % Ethanol	1 h
40 % Ethanol	1 h
50 % Ethanol	1 h
60 % Ethanol	1 h
70 % Ethanol	2 h
85 % Ethanol	over-night
95 % Ethanol	4 h
100 % Ethanol	4 h
100 % Ethanol	over-night or until further procedure

To prevent RNA degradation, samples were kept at 4°C on ice until 85 % ethanol was reached.

9.5.2 Staining and embedding of samples

Samples were stained with eosin (0.1% Eosin Y in 100 % ethanol). Stained samples were placed in an automated embedding system (ASP300 tissue processor, Leica), which exchanges the ethanol with liquid paraffin (Paraplast Plus, McCormick). Samples were placed separately in plastic molds, orientated for longitudinal sectioning, covered with paraffin and solidified. Embedded samples were stored at 4°C.

9.5.3 Sectioning of embedded samples

Sectioning was performed on a rotary microtome (Leitz 1512) at a thickness of 7 μm. Ribbons of sections were laid on adhesive glass slides (Superfrost ® Plus, Menzel) covered with water and placed on a heating plate at about 40°C. When ribbons were stretched out, the water was removed and sections were dried over the heating plate over-night.

9.5.4 Synthesis of ribo-probes

Templates for probe synthesis were obtained by PCR on Col-0 cDNA. T7-RNA polymerase promoter sequence (5´-TAATACGACTCACTATAGGG-3´) was added to the reverse primer (for primer sequences and probe length see Appendix A, Tab. A.6). Probe synthesis was performed at 37°C for 60 min using T7-RNA Polymerase (Roche) in 25 μl of the following reaction mix: 1x RNA Polymerase buffer (Roche) containing 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 0.4 mM DIG-UTP, 50 ng PCR product, 40 U RNase Inhibitor (Roche), 20 U T7 RNA-Polymerase.

To stop the reaction, 75 μl 1X MS (10 mMTris-HCI pH 7.5, 10 mM, 50 mMNaCl), 2 μl tRNA 100 mg/ml, 1 μl DNase (RNase free, 10 U/μl) were added and incubated for 10 min at 37°. RNA was precipitated with 100 % Ethanol, washed once with 70 % Ethanol, air dried and resolved in 50 μl Tris/EDTA (TE) buffer.

9.5.5 Tissue pretreatment

Slides were placed in stainless steel racks and passed trough the following solutions, numbers indicate when fresh solutions where used:

Histoclear (national diagnostics) 1	10	min	
Histoclear 2	10	min	
100 % Ethanol 1	1	min	
100 % Ethanol 2	30	sec	
95 % Ethanol	30	sec	
85 % Ethanol, 0.85 % NaCl	30	sec	
50 % Ethanol, 0.85 % NaCl	30	sec	
30 % Ethanol, 0.85 % NaCl	30	sec	
0.85 % NaCl	2	min	
PBS 1	2	min	
Proteinase K (1 μg/ml in 100 mM Tris pH 8.0, 50 mM EDTA)	30	min	37°
0.2 %glycine in PBS	2	min	
PBS 1	2	min	
PBS 2	2	min	
4 % paraformaldehyde in PBS	10	min	
PBS 2	2	min	
PBS 3	2	min	
0.5 % acetic anhydride in 0.1 M triethanolamine pH 8.0	10	min	
PBS 3	2	min	
0.85 % NaCl	2	min	

Samples were dehydrated by passing them back through the ethanol series (from 30 % ethanol to 100 % ethanol) and kept at 4°C with ethanol covering the bottom of the slides until further procedure.

9.5.6 Hybridisation

Hybridisation buffer for 48 slides was prepared according to the following recipe:

10x salts (3 M NaCl, 0.1 M Tris-HCl pH 6.8, 0.1 M $NaPO_4$, 50 mM EDTA)	240 µl
deionized formamide	960 µl
tRNA 100 mg/ml	24 µl
50x Denhardt´s (Sigma)	48 µl
H_2O (sterile)	160 µl
50 % dextran sulphate	480 µl
final volume	1920 µl

Slides were dried under a fume hood and sections were marked with a Pap pen (marking pen that provides a thin hydrophobic barrier) to keep the hybridisation buffer in the area of the sections later on. Probes were heated to 95°C for 2 min and immediately cooled on ice afterwards. For one slide, 4 µl of probe were mixed with 4 µl deionized formamide and 32 µl of hybridisation buffer and added onto the sections. Slides were covered with a coverslip and hybridised overnight in a humid atmosphere (a sealed box with water-soaken paper at the bottom) at 50°C.

9.5.7 Washing and antibody staining

Washing was performed in 0.1x saline sodium-citrate buffer (SSC) (1x SSC: (0.15 M NaCl, 15 mM Na_3Citrate)). Slides were placed in stainless steal racks again and washed under gentle agitation for 5 min to make the coverslips fall off. Slides were then washed at 50°C for 30 min, 2x 45 min and 1 h and 5 min in PBS at room temperature. For the antibody staining, slides were incubated for 5 min in Buffer 1 (100 mM Tris-HCl, 150 mM NaCl, pH 7.5). All following incubations were performed in square petri dishes under soft agitation with the buffers completely covering the slides. Slides were transferred to fresh petri dishes after each step. Slides were incubated in Buffer 2 (0.5 % (w/v) blocking reagent (Roche) in Buffer 1) for 30 min, Buffer 3 (1 % BSA, 0.3 % Triton X-100 in Buffer 1) for 30 min, Buffer 4 (Anti-digoxigenin-AP FAB (Roche) 1:3000 in Buffer 3) for 1.5 h, Buffer 3 4x 20 min, Buffer 1 and Buffer 5 (100 mM Tris pH 9.5, 100 mM NaCl, 50 mM

$MgCl_2$) for 5 min each. For the color reaction, slides were incubated in Buffer 6 (150 μg/ml nitro blue tetrazolium (NBT), 75 μg/ml 5-Bromo-4-chloro-3-indolyl phosphate (BCIP), 24 μg/ml levamisole) in the dark.

9.5.8 Washing and counterstaining

The duration of the color reaction was varied according to signal strength from 17 h to 41 h. To stop the reaction, slides were placed back in stainless steel racks and washed in the following solutions:

H_2O (sterile)	3 min
H_2O (sterile)	30 sec
70 % Ethanol	30 sec
95 % Ethanol	30 sec
100 % Ethanol	30 sec
95 % Ethanol	30 sec
70 % Ethanol	30 sec
H_2O (sterile)	30 sec
H_2O (sterile)	30 sec

Slides were then dried under a fume hood. 2 drops 50 % glycerol were added to each slide and coverslips were lowered onto them. Slides were left over night in fume hood and then viewed and photographed using DM RB light microscope (Leica).

9.6 Global expression profiling

To determine differentially expressed genes between Col-0 and candidate gene knock out lines, genome-wide expression data were obtained using AGRONOMICS1 Tiling Array (Affymetrix®). In contrast to the ATH1 expression array, which was the previous standard array for Arabidopsis gene expression analysis and contained probes for about 22 000 genes, this array contains genome-wide tiled probes, enabling expression analysis for over 29 000 genes when TAIR9 genome release is applied (Rehrauer et al., 2010).

9.6.1 RNA preparation and quality control

Total RNA from plant material was obtained as described in 4.11. Further processing was done at the Max Planck Genome Centre Cologne. RNA was further purified and concentrated using MinElute PCR Purification Kit (Qiagen). Quality of RNA was accessed with Bionanalyzer 2100 (Agilent). Three biological replicates were processed for each genotype.

9.6.2 cRNA Synthesis and array hybridisation

The cRNA Synthesis and array hybridisation were done at the Max Planck Genome Centre Cologne as described in Rehrauer et al. (2010):

Total RNA samples (300 ng) were reverse transcribed into double-stranded cDNA and then *in vitro* transcribed in the presence of biotin-labelled nucleotides using the MessageAmp™II-Biotin Enhanced Single Round aRNA Amplification Kit (Ambion; A-M1791) including poly(A) controls as recommended by Affymetrix. The quality and quantity of the biotinylated cRNA were determined using NanoDrop ND 1000 and Bioanalyzer 2100. Biotin-labelled cRNA samples (15 μg) were fragmented randomly to 35 to 200 bp at 94°C in Fragmentation buffer (Affymetrix).

Biotin-labelled cRNA samples were mixed in 300 μl of Hybridisation Mix (Affymetrix; 900720) containing Hybridisation Controls and Control Oligonucleotide B2 (Affymetrix; 900454). Samples were hybridised onto Affymetrix AGRONOMICS1 Arabidopsis tiling arrays for 16 h at 45°C. Arrays were then washed using an Affymetrix Fluidics Station 450 using the FS450_0004 protocol. An Affymetrix GeneChip Scanner 3000 was used to measure the fluorescence intensity emitted by the labelled target.

9.6.3 Normalisation of expression values and annotation to genome

All data processing was done using R Version 2.11.1. The aroma.affymetrix package (Bengtsson et al., 2008) was used together with the scripts agronomicsTools01.r to perform RMA normalisation and calculation of mean values over all probes per gene. To assign probes to genes, a CDF file was created according to aroma.affymetrix instructions using AGRONOMICS1_At_TAIRG R-package (Version 13.0.0, TAIR9 annotation)

obtained from the Brainarray project webpage (Dai et al., 2005).

9.6.4 Determination of differentially expressed genes

Differentially expressed genes were determined using the bioconductor (www.bioconductor.org) package RankProd (Hong et al., 2006). Genes with false discovery rate corrected p-values below 0.05 were considered as differentially expressed.

10.1 Genomic structure and sequence information of DPA4

The *DPA4* gene is located on the 5th chromosome of Arabidopsis (AGI locus code At5g-06250) in reverse orientation. The predicted transcript consists of two exons and one intron. Two splice variants are predicted in the gene model, varying in size of the first exon (Fig. 10.1). Predicted DPA4 proteins contain 267 (At5g06250.1) and 282 amino acids (At5g0650.2), both coding for a putative B3 domain containing transcription factor. At5g0650.2 is the representative gene model used in the data base of The Arabidopsis Information Resource (TAIR) and was also found to be the product amplified from cDNA in our work (data not shown). A Blast search on At5g06250 revealed highest sequence homology to At3g11580. At5g06250.1 shares 69.0% of its nucleotides with At3g11580, At5g0650.2 displays 69.5% identity in nucleotides with At3g11580 on mRNA level. Comparison on amino acid level displays 58.4% identity for At5g06250.1 and 58.5% identity for At5g0650.2. Both isoforms of DPA4 and At3g11580 contain the repressive motif described for B3 transcription factors (L/V RLFGV N/D M/L/V) in the variety VRLFGVNL (Ikeda and Ohme-Takagi, 2009).

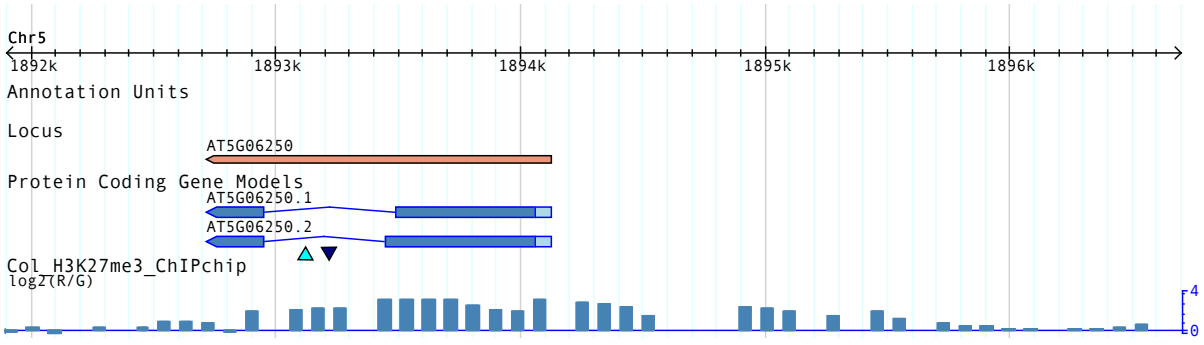


Figure 10.1: **Characteristics of *DPA4* locus.** *DPA4* locus with associated gene models, T-DNA insertions used in this study (*dpa4_1*-light blue triangle, *dpa4_2*- dark blue triangle) and distribution of the histone mark H3K27me3.

10.1.1 Functional predictions for DPA4 according to sequence information

Due to its sequence similarity, DPA4 is predicted to be a member of the B3 superfamily and a RAV transcription factor (Romanel et al., 2009). As the NGA proteins, it only contains a B3 domain and no AP2 domain (see 8.4.1). The affiliation to this group of transcription factors does not directly allow predictions about putative functions of DPA4 or target genes, but suggests that targets might contain the CACCTG sequence in their promoters. Since already characterised members of the RAV family are involved in flowering time regulation and leaf and gynoecium development, DPA4 could regulate genes involved in these processes as well.

The biological function of the closest homolog to DPA4, At3g11580 is not yet discovered, therefore no conclusions can be drawn from this similarity. Two T-DNA insertion lines in the At3g11580 locus were grown and selected for homozygous plants, but no abnormal developmental phenotype could be observed in those lines.

10.2 Phenotype of DPA4 associated T-DNA insertion lines

10.2.1 Reduction of *DPA4* transcript

Two T-DNA insertion lines associated with the DPA4 locus were analysed, *dpa4_1* and *dpa4_2* (lines I10 and D3 from Part I). For both lines, the T-DNA is inserted in the intron region (Fig. 10.1) but in opposite orientation. Although *dpa4_2* was supposed to be located in an exon according to records of the SALK institute, it was shown to be located in the intron region by sequencing (data not shown).

Expression of DPA4 is strongly reduced in *dpa4_2* (Fig. 10.2). In *dpa4_1* DPA4 expression seems to be increased as measured by qRT-RCR. The primers used for this experiment are designed to amplify a part of the first exon, the reverse primer spans the exon-exon border to avoid amplification of genomic DNA. Two explanations are possible to explain the high m-RNA levels measured in *dpa4_1*: either the intron inserted T-DNA is transcribed and spliced out with the rest of the intron or only a partial transcript is made and the reverse primer only hybridises to a the part belonging to exon 1. There-

fore, a conventional PCR was performed to detect the full-length transcript in both lines. No full-length transcript could be observed in *dpa4_2* (Fig. 10.2 B), in *dpa4_1*, very low amounts are visible. This indicates that the insertion in *dpa4_2* plants is able to prevent transcription of the complete locus, whereas transcription can read over the insertion in *dpa4_1*, even though the efficiency seems to be very low. To explain the high levels of mRNA detected in *dpa4_1* by qRT-PCR it can be assumed, that a partial transcript is made.

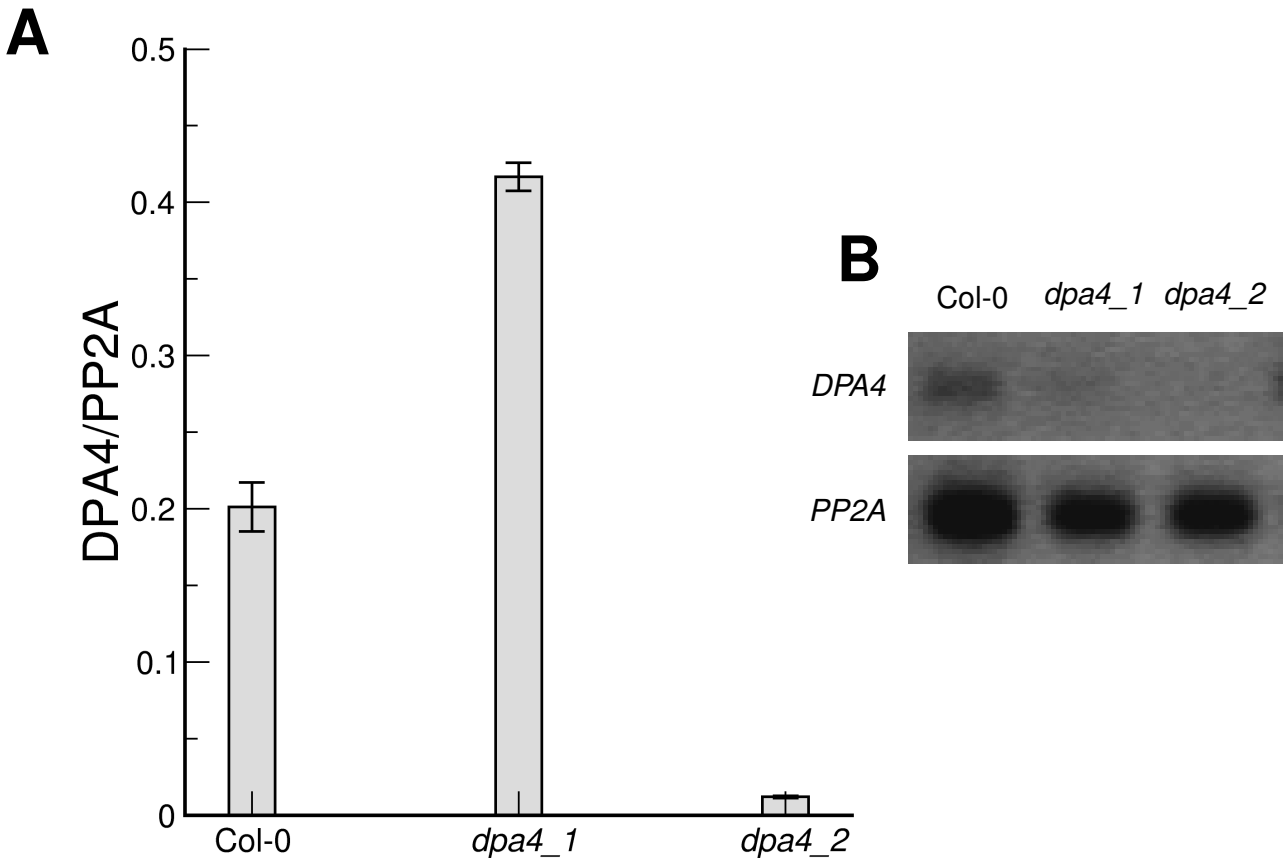


Figure 10.2: **Expression of *DPA4* in associated T-DNA insertion lines A**: Expression of *DPA4* measured by qRT-PCR. RNA of 11 day old seedlings harvested at ZT 4 was used. Error bars indicate standard deviation. Results are based on a single experiment, similar results for *dpa4_2* were observed several times and at different developmental stages. Expression of *dpa4_1* was equal to Col-0 in some experiments. **B**: Amplification of nearly full-length transcript of *DPA4*, primers used were those for cloning full-length *DPA4* coding region (Tab. A.4). RNA samples were same as in A.

10.2.2 Leaf shape and floral organ size in *dpa4*

Both *dpa4_1* and *dpa4_2* display enhanced leaf margin serration (Fig. 10.3 A) and enlarged petal size (Fig. 10.3 C). No significant difference in size could be observed for other floral organs, although a tendency for slightly bigger sepals in *dpa4_2* plants was observed in a preliminary measurement of a few flowers. To test whether this difference is significant, the sample size has to be enlarged. Serrations are stronger in *dpa4_2* than in *dpa4_1*. The weaker phenotype of *dpa4_1* can be explained by the residual amount of full-length transcript in this line (see 10.2.1). Therefore, the following experiments were carried out with *dpa4_2* plants when possible.

Detailed analysis of leaf shape

Serration in *dpa4* plants is mainly increased by a deeper sinus. At a first glance the number of teeth seems to be increased in *dpa4*, but it is similar compared to wild type when taking a close look to the wild type leaf and taking also very small teeth into account. As in wild type, serrations increase in later produced leaves (Fig. 10.3 A) of *dpa4* plants and also the differences compared to the wild type become more severe. Rosette diameters are on average the same in *dpa4* and Col-0 plants (data not shown), suggesting no alteration in overall length of rosette leaves. Measurements of leaf surface for 6^{th}, 7^{th}, 8^{th} and 9^{th} leaf of 7 plants did not reveal any significant difference in overall leaf surface between Col-0 plants and *dpa4_2* plants (data not shown). Nevertheless the shape of *dpa4* leaves apart from serrations is also slightly different compared to wild type: *dpa4* leaves are narrower at the proximal part and more expanded at the distal part than Col-0 leaves (Fig. 10.3 B).

Detailed analysis of petal shape

Petals are significantly longer in both *dpa4* lines to a similar extent (petals in both *dpa4* lines are about 0.7 mm longer than in Col-0, Fig. 10.3 C). Petals also seem to be slightly altered in shape: in relation to the visible distal part, the proximal part is slimmer than in Col-0 plants (Fig. 10.3 D). Preliminary comparisons of 4 pairs of flowers indicate, that the surface of petals is increased in both *dpa4* lines (data not shown).

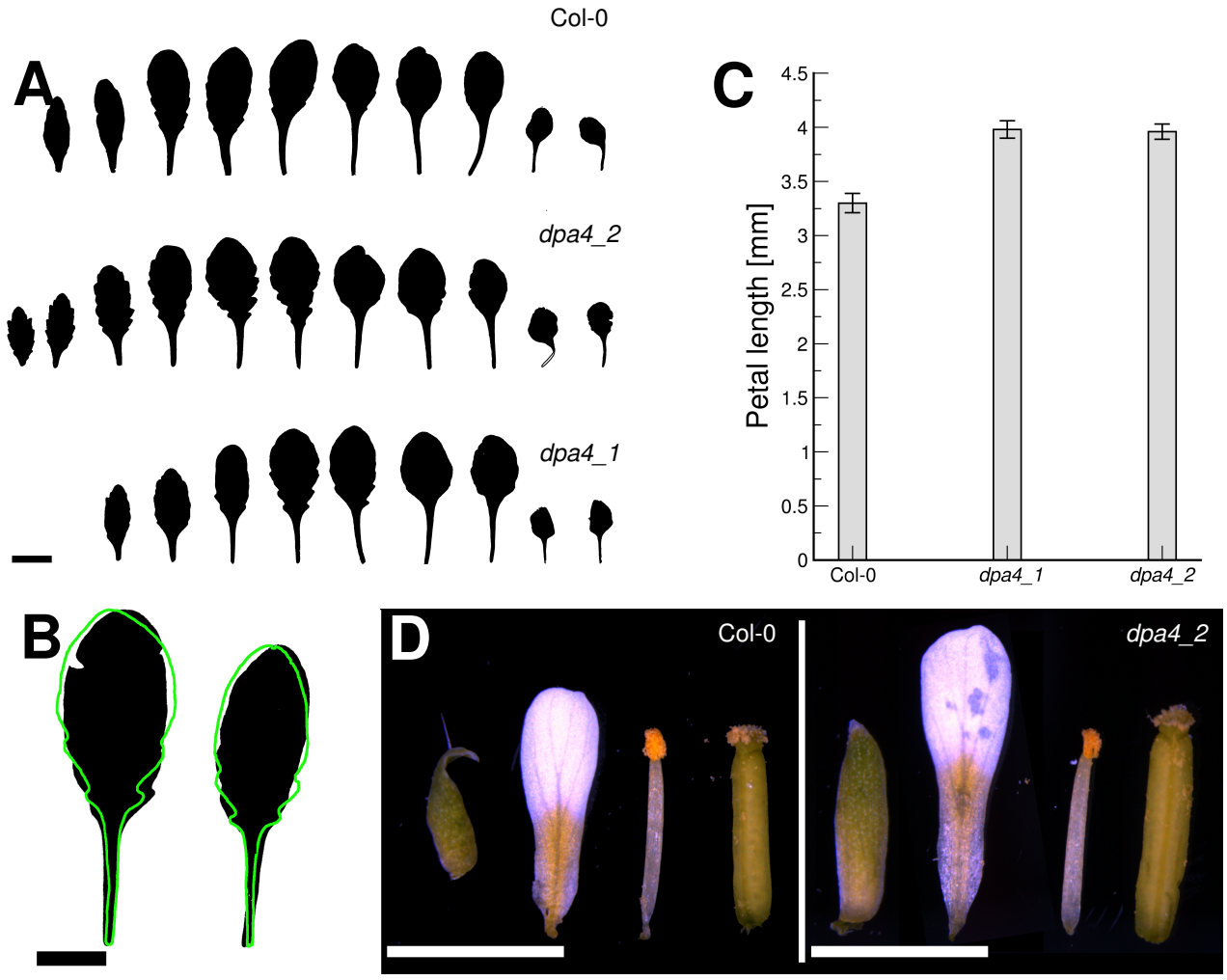


Figure 10.3: **Phenotype of *dpa4*.** **A**: Leaf series of Col-0, *dpa4_2* and *dpa4_1*, 21 day old plants grown in LD. **B**: Overlay of Col-0 leaf shape (black) with *dpa4_2* leaf shape (green line), 6th leaves of approximate similar sizes from 28 day old plants grown in LD. **C**: Petal length of flowers stage 14 in Col-0, *dpa4_2* and *dpa4_1*, plants grown in SD. Error bars indicate standard error of the mean. **D**: Dissected floral organs of Col-0 and *dpa4_2* stage 14 flowers from SD grown plants. Scale bars: 1 cm in A and B, 2 mm in C .

10.2.3 Epidermal cell shape and cell number in leaves and petals of *dpa4*

While in leaves the epidermal cell shape is not altered in *dpa4* plants, the epidermis cells of petals seem to be slightly more copped than Col-0 cells (Fig. 10.4). This result was only observed for few petals so far and has to be verified in future experiments. Counting of cells per area for both petals and leaves revealed no alteration in cell number per area, which means that in leaves, where surface area is not altered, cell numbers are similar to wild type, while petals of *dpa4* plants contain more cells than wild type plants, indicating

an either faster proliferation rate or a longer proliferation period. Margin cells in *dpa4* plants seem to elongate and develop normally, except for the increase in sinus depth and the associated stronger bending of the leaf margin cells no difference in cell shape is visible in serrations of *dpa4* (Fig. 10.4 C and D).

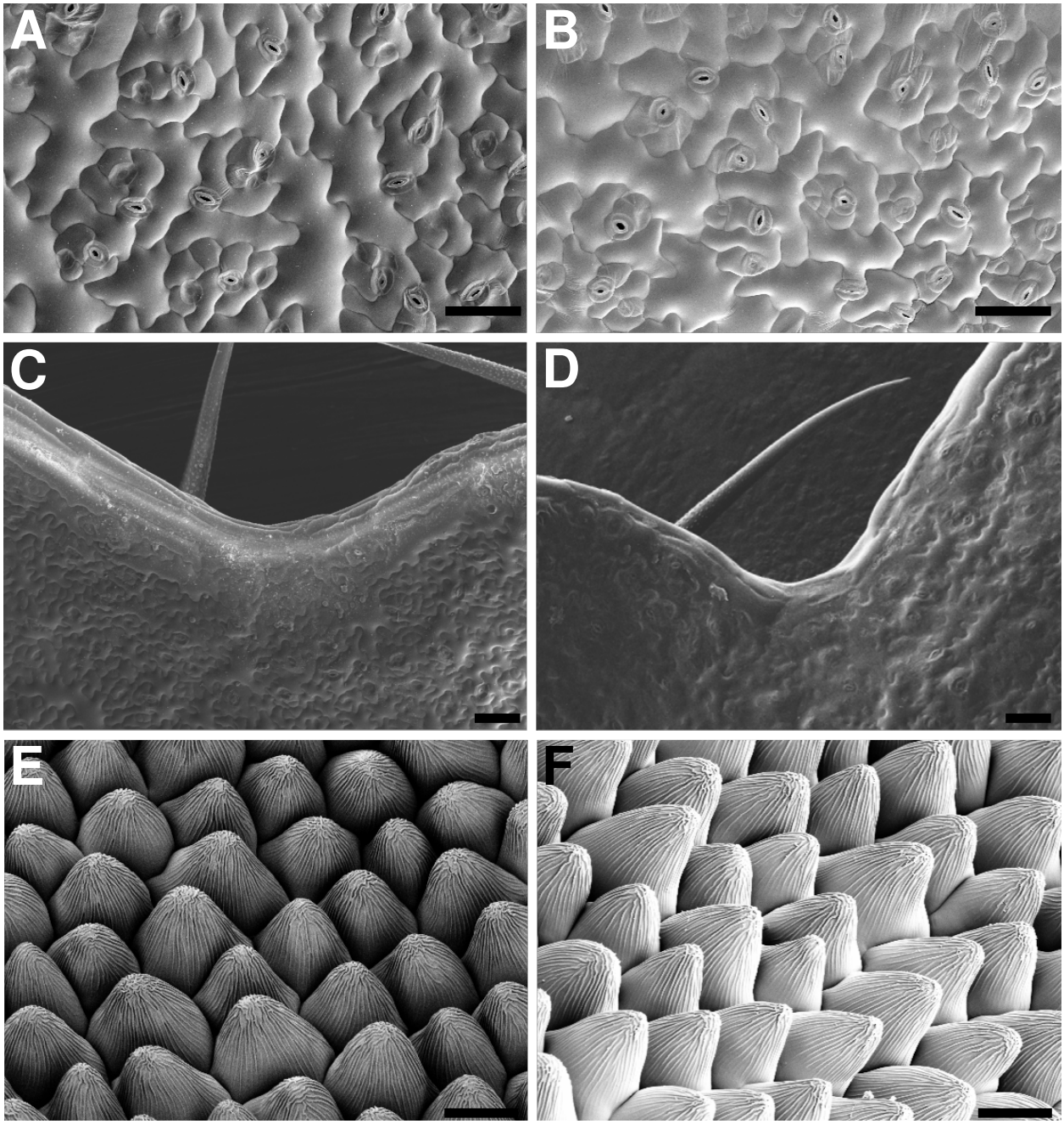


Figure 10.4: **Epidermal cell shape in *dpa4*.** Scanning electron microscopy images of epidermal cells: Col-0 young leaf of 5 week old plant in SD (A), *dpa4_2* young leaf of 5 week old plants in SD (B), abaxial site of 7th leaf of 3 week old plant in LD for Col-0 (C) and *dpa4_1* (D), petals of open flowers for Col-0 (E) and *dpa4_2* (F). Scale bars: 40 μm in A-D, 10 μm in E and F.

10.3 Phenotype of *35S::DPA4* lines

Only very few of the *35S::DPA4* primary transformants germinated. Those seed were already selected for GFP signal in the seed coat and should therefore all contain the transgene. 3 plants with a similar abnormal phenotype were found in the T_1 generation, all of them had smooth, narrow leaves without any serrations. Two plants flowered, although later than wild type and with a low yield of seed. The third plant was sown later and is still growing, but the phenotype seems to be very severe in this plant and it is not sure that it will flower or even survive.

Analysis of the segregation ratio among the seed of the two first plants revealed a ratio of 2.2 and 2.1 between GFP carrying seeds and normal seeds. This result is still in the range to be considered a 3:1 segregation indication a single insertion locus. Single locus insertions are preferred because they are more likely to result in stable expression levels. Therefore, the GFP selected seeds of these two plants were used for further analysis (lines *35S::DPA4_1* and *35S::DPA4_2*).

Again some T_2 seeds did not germinate, but the number of total seeds (30 seeds sown, 24 germinated) was to low to draw conclusions from that fact. To quantitatively test for the germination efficiency a germination assay with a higher number of seed will be performed in future for homozygous seeds.

All recovered seedlings of the T_2 generation displayed narrow leaves without any serrations (Fig. 10.5 A and B). Older seedlings also showed a twisted leaf phenotype, leaves were turned like a left handed spiral (Fig. 10.5 C). *35S::DPA4* plants flower at least 10 days later than wild type and display a strong variance in flowering time (Fig. 10.5 D). Exact flowering time in terms of total leaf number will be determined in next generation, but it is already visible that *35S::DPA4* plants also produce more leaves until bolting. Plants with a very strong abnormal phenotype are also smaller (Fig. 10.5 D). Plants with milder phenotypic changes are similar to wild type in size during vegetative development (Fig. 10.5 A and B), but rosette leaves do not expand to the same degree as wild type leaves after bolting.

Inflorescences of *35S::DPA4* plants are fasciated and some flowers display reduced numbers of petals (Fig. 10.5 E and F). Siliques are abnormally shaped in the overexpressors as well, the replum is broader than in wild type and the valves seem to grow out at

the top in an abnormal way, causing the siliques to become broader instead of narrower towards the end (Fig. 10.5 G and H).

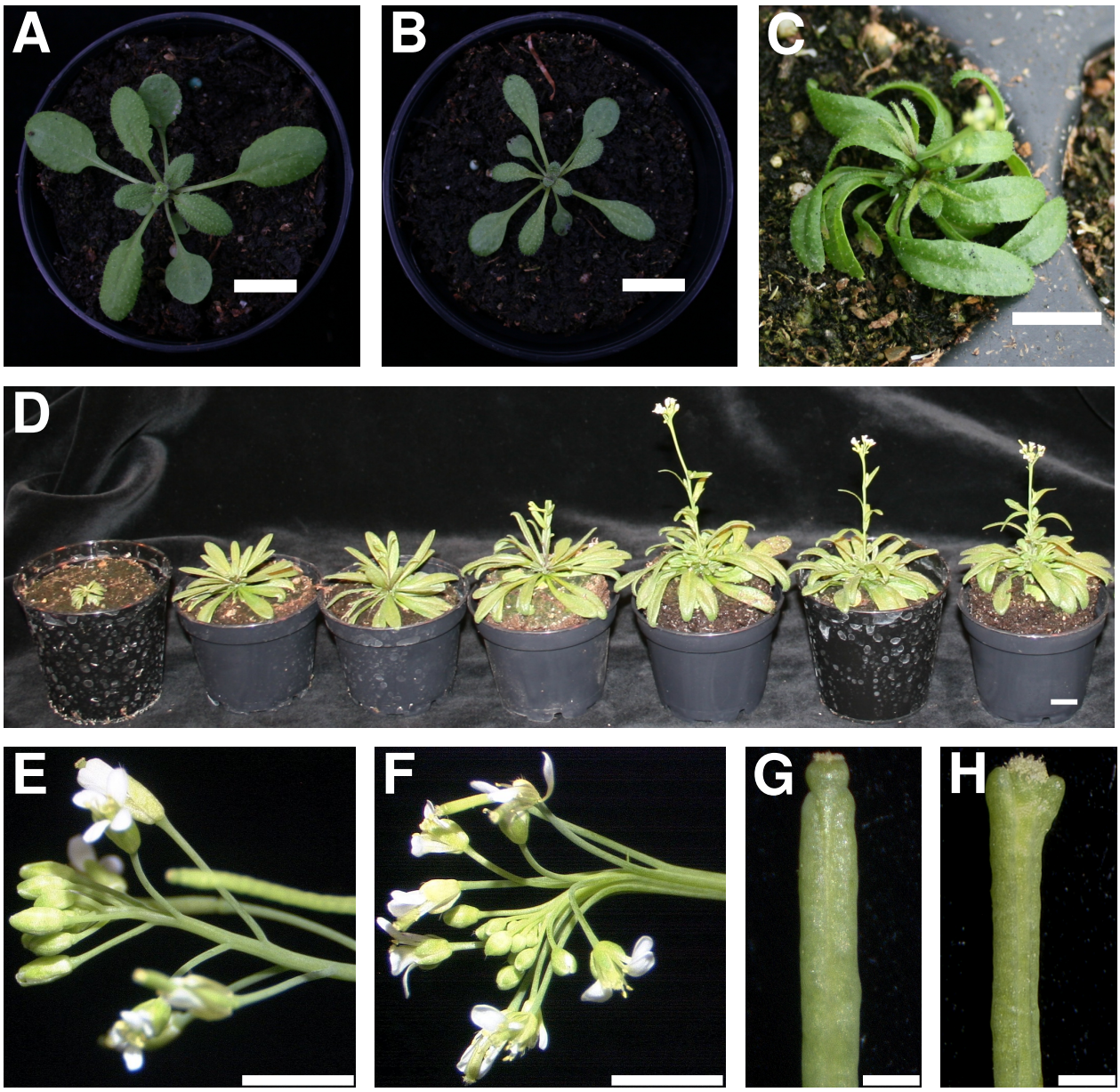


Figure 10.5: **Phenotype of *35S::DPA4* plants in LD. A**: Rosette of 21 day old Col-0 plant. **B**: Rosette of 21 day old *35S::DPA4_1* plant (T_2 generation). **C**: Twisted rosette leaves in 7 week old *35S::DPA4_2* plant (T_1 generation). **D**: Phenotypic variance in T_2 generation of *35S::DPA4_1* plants, all plants are 5.5 weeks old. Col-0 control plants are already setting seeds at this stage (not in the picture). **E**: Inflorescence of 5.5 week old Col-0 plant. **F**: Inflorescence of 5.5 week old *35S::DPA4_1* plant, the stem is strongly fasciated towards the top. **G**: Detailed view on a Silique of 5.5 week old Col-0 plant. **H**: Detailed view on a Silique of 5.5 week old *35S::DPA4_1* plant. Scale bars: 1 cm in A - D, 0.25 cm in E and F, 0.1 cm in G and H.

10.4 Expression Pattern of *DPA4*

Since *DPA4* was a gene stably assigned to the apex/flower cluster in the analysis performed in Part I, *DPA4* should be expressed mainly in the shoot apex and in flowers. Visualisation of the Developmental Series data with the Arabidopsis eFP Browser (www.bar.utoronto.ca) shows strongest expression of *DPA4* in the inflorescence shoot apex and siliques containing stage 3 seeds. Strong expression is also observed in transition / vegetative apices and flowers at stage 9 (Fig. 10.6 A). Expression in leaves is about half as strong as the strongest expression in shoot apices.

The phenotype of *dpa4* suggests a role of DPA4 not only in flower but also in leaf development. Together with the expression pattern, this leads to the question, whether DPA4 controls development rather at early stages in the shoot apex or at later stages in the developing tissue. To answer this question, expression analysis was performed using qRT-PCR and *in situ* hybridisation.

10.4.1 Expression in different tissues as revealed by qPCR methods

qRT-PCR was performed to determine expression of *DPA4* in different tissues at different stages to confirm the electronic data and to determine the tissue best suited for further *in situ* hybridisation and genome-wide expression analysis. The observed expression values for *DPA4* are in accordance with the expression values provided by the eFP browser: *DPA4* expression is always higher in apex enriched samples than in the corresponding leaf samples. In the early time point (11 day old seedlings), expression in LD is in general higher than in SD. Strongest expression of *DPA4* is observed in 28 day old SD grown plants. In accordance with this result the following genome-wide expression experiment (to determine putative DPA4 target genes) was done on apex enriched samples of 28 day old plants grown in SD (Fig. 10.6 B).

10.4.2 Detailed spatial expression pattern as revealed by *in situ* hybridisation

The first probe for *in situ* hybridisation which gave significant signal (probe 2) was designed to hybridise to the first exon of *DPA4* and a few bases of the second intron. This

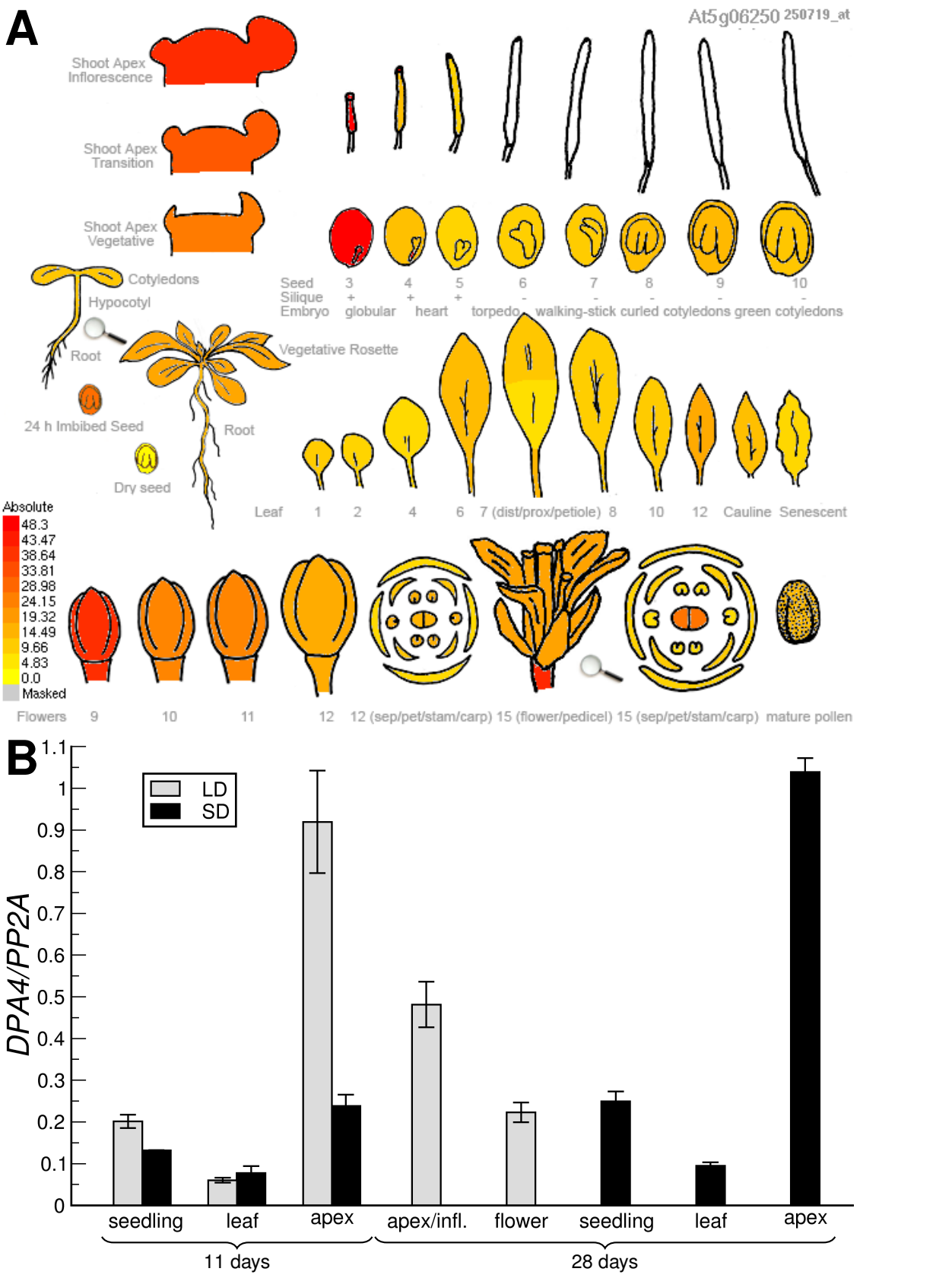


Figure 10.6: **Spatial expression pattern of *DPA4* at different developmental stages. A**: Arabidopsis eFP Browser view of MAS 5.0 normalised absolute expression data of the "Developmental series" for *DPA4*. Expression values are indicated by a colour code. **B**: Expression of *DPA4* at different tissues and developmental stages in LD and SD measured by qRT-PCR. Error bars indicate standard deviations.

probe also gave a signal in *dpa4_2* samples. To test whether this was due to cross hybridisation to another transcript or because of a partial transcript also made in *dpa4_2* (as already observed for *dpa4_1*, see 10.2.1), a new probe (probe 3) was designed spanning the exon-exon border in the middle of the probe. This probe gave the same expression pattern as probe 2 in wild type but did not show signal in *dpa4_2* apices (Fig. 10.7). Probe 3 was used for all further experiment. However, since probe 2 also seems to indicate the true expression pattern of *DPA4* in wild type, not all stages were repeated with probe 3.

The *in situ* hybridisation experiments show expression of *DPA4* in the shoot apex, restricted to the areas of primordia formation. In vegetative and transition apices, expression is observed at the emerging boundary between the shoot apex and the leaf primordium. *DPA4* expression persist in this area as the leaf develops (Fig. 10.7 A and Fig. 10.8 A and B). In inflorescence apices, *DPA4* is expressed in the boundary between the inflorescence meristem and the emerging flowers, in floral meristems *DPA4* expression marks the boundaries between emerging floral organs (Fig. 10.8 C and D). In flowers *DPA4* expression persist between organs, also between the two fused carpels (Fig. 10.8 E). Thus, *DPA4* is expressed wherever organs separate from the shoot apical meristem or from each other. In addition to that, *DPA4* expression could be observed in the leaf sinus, suggesting a role of *DPA4* in early stages during organ initiation as well as during early development in young organs (Fig. 10.8 F and G).

DPA4 expression was never observed in the center of the meristem and only becomes visible where organ or lateral meristem emergence is already initiated. Furthermore, expression seems to occur in different layers in the SAM compared to leaves and floral meristems: in the SAM *DPA4* expression is excluded at least from L1 and L2 (Fig. 10.8 A,B and C), while in floral meristems and leaves expression occurs directly in the outer layers.

10.5 Expression of *DPA4* in PcG mutants

The very restricted spatial expression pattern of *DPA4* in organ boundaries suggest a strong regulation of *DPA4* expression. The *DPA4* locus including the promoter region is widely covered with the chromatin repression mark H3K27me3, suggesting a regulation of *DPA4* expression by the PcG pathway (Fig. 10.1). Nevertheless, *DPA4* was not identi-

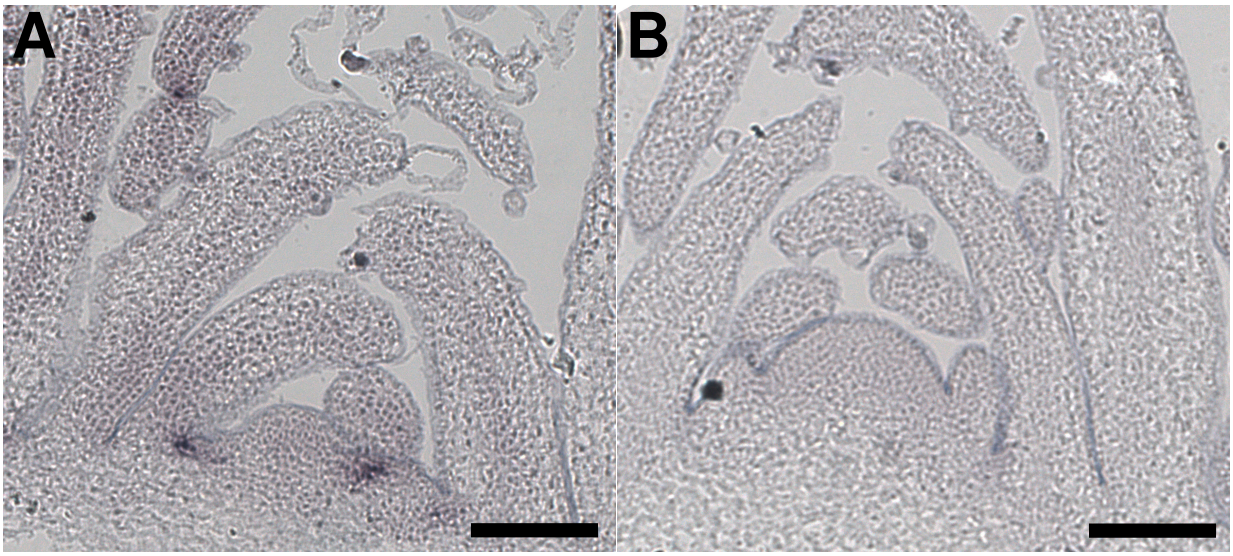


Figure 10.7: **Verification of specificity for *in situ* probe 3.** In situ hybridisation of Col-0 (A) and *dpa4_2* apices (b) of 28 day old plants grown in SD with probe 3. Scale bars: 100 μm.

fied as an upregulated gene in a genome-wide expression analysis of the PcG knock out mutants *clf* and *clf/swn* (personal communicating by Prof. Justin Goodrich). However, altered *DPA4* expression was observed in a qRT-PCR analysis on *clf/swn* double mutants (Fig. 10.9) at 11 and 22 days. In the single *clf* mutant *DPA4* expression is slightly lower compared to wild type, but this deviation could also be in the range of natural changes in expression. In this context it has to be noted that *clf* plants are early flowering and therefore may not exactly be in the same developmental stage as Col-0 plants, even at 11 days. The result suggests that either SWN alone can maintain the repression of *DPA4* or that both components are not needed for transcriptional regulation of *DPA4*. An up to 6 times upregulation of *DPA4* in the double mutants of the partially redundant components CLF and SWN in 22 day old plants indicates that *DPA4* is indeed regulated by the PcG system or at least by some mechanisms affected in the *clf/swn* double mutant that shows a very pleiotropic growth phenotype (see 1.2.1).

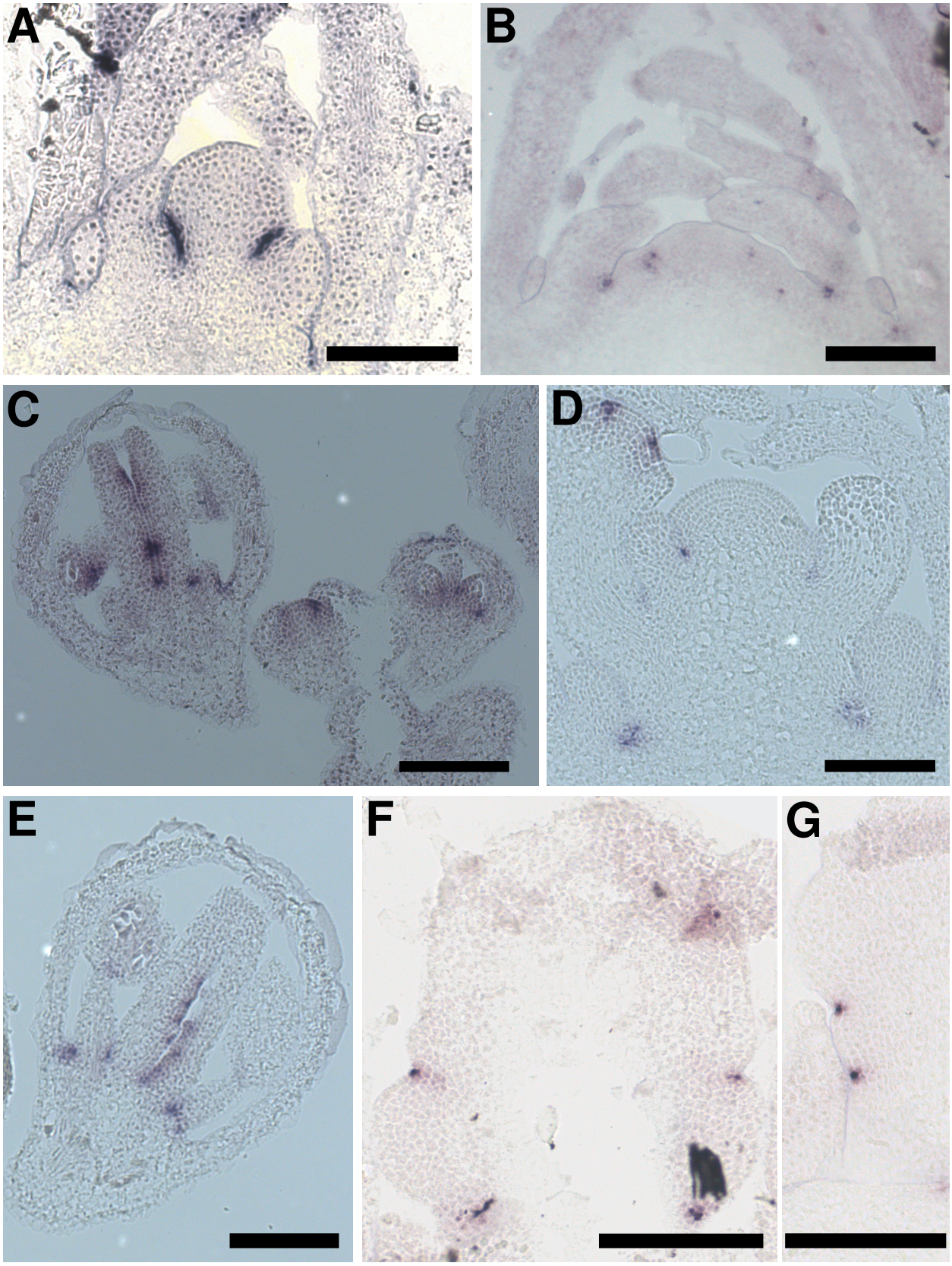


Figure 10.8: **Spatial Expression pattern of *DPA4*.** Expression of *DPA4* in transition apex of 11 day old plant in LD (A), in vegetative apex of 35 day old SD grown plant (B), in a flowering apex of 28 day old plant in LD (C), in an inflorescence apex of a plant that has been shifted to LD for 7 days after 28 days in SD (D), in a young flower (E) and in young leaves (F and G). Scale bars: 50 μm in A, 100 μm in B-G.

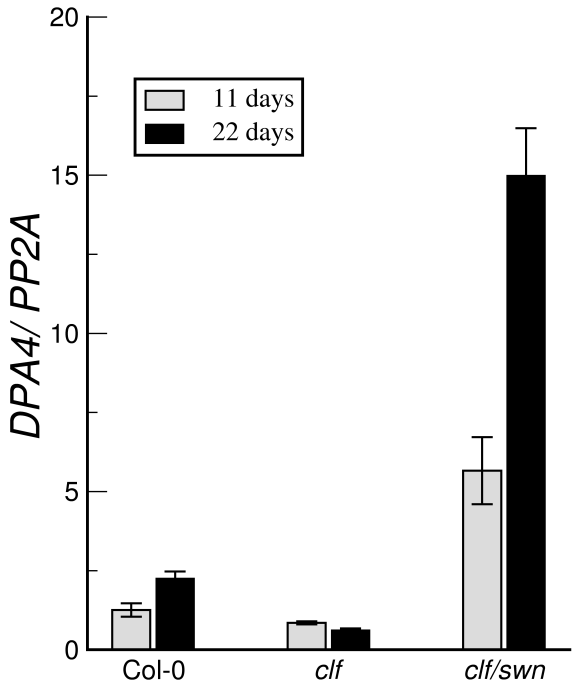


Figure 10.9: ***DPA4* expression in PcG mutants measured by qRT-PCR.** Samples were obtained from whole seedlings at ZT9 in LD. Data are based on a single experiment. Error bars indicate standard deviations.

10.6 Transcript profile in *dpa4* apex enriched tissue

Since DPA4 is a putative transcription factor, it is expected to regulate transcription of several target genes and thereby perform its role in leaf and flower development. Therefore, a transcriptional profiling of *dpa4* in comparison to wild type tissue was performed to determine putative targets of DPA4. As tissue type, "apex" tissue (harvested as described in 4.11) of 28 day old plants grown in SD conditions was chosen, since this tissue shows strong expression of *DPA4*. Furthermore, the leaf serration phenotype becomes visible at that stage in SD conditions (data not shown). Triplicates of Col-0 and *dpa4* samples were hybridised to AGRONOMICS1 Tiling Array to perform a genome-wide transcriptional profiling. Determination of differentially expressed genes using rank product statistics revealed 16 significantly upregulated (Up) and 61 significantly downregulated (Down) genes in *dpa4* compared to wild type. Cut off for false discovery rate corrected p-values was 0.05.

As DPA4 carries the repressive motif, one would rather expect more genes to be upregulated in the mutant than downregulated, but the effect on the downregulated genes

could also be indirect through some upregulated genes.

10.6.1 Hexameric motifs in the promoters of differentially expressed genes

If DPA4 would act as a transcriptional regulator in the same way as other RAV proteins, it should bind to the binding sequence of the RAV B3 domain (CACCTG). Therefore, the promoters of the differentially expressed genes were probed for the existence of this motif using "Motif Analysis" (TAIR, see 9.1). 8 of the upregulated and 24 of the downregulated genes contain this motif in their promoter region. Promoter was defined as 3000 bp upstream of the transcriptional start site. Both numbers are not statistically different from what is expected in a random gene set. Interestingly, the AuxRE TGTCTC was found in 57 of the 61 downregulated genes, which is a significant overrepresentation ($p = 0.00198$ assuming binomial distribution). The element was also found in 11 of the upregulated genes, but this number does not correspond to a significant overrepresentation of the motif.

There were 241 overrepresented motifs in the promoters of the upregulated genes and 807 in the promoters of downregulated genes ($p < 0.05$ assuming binomial distribution). This enormous number of motifs prohibits a detailed analysis for a hint to explain the genes responsible for the *dpa4* phenotype. Since the number of differentially expressed genes is quite small, especially in case of the downregulated genes, a high percentage can be verified by qRT-PCR. This will be done in future and after the exclusion of putative false positive genes the motif analysis will be repeated.

10.6.2 Characteristics of genes differentially expressed in *dpa4*

Analysis of the functional characterisation state of the differentially expressed genes in *dpa4* revealed that only one of the genes is experimentally characterised so far, most functions assigned were just predicted by sequence similarity. In addition, 10 upregulated and 49 downregulated genes are completely unknown (Tab. 10.1). One might argue that the usage of a whole genome tiling array elevates the number of unknown genes, since the additional genes compared to the former standard array were just recently annotated. 24 of the downregulated and 3 of the upregulated genes were present on the ATH1 array,

but those did not contain a lower fraction of transposons or completely unknown genes compared to the whole data set. In fact, this observation reveals that whole genome tilling array resulted in more than twice the number of differentially expressed genes in this case.

Table 10.1: **Summary of unknown differentially expressed genes in *dpa4*.** Differentially expressed genes were determined using RankProd at cut off p < 0.05. Information about genes was received from TAIR.

Type	Number in Up	Number in Down
Unknown protein/ no blast hit in other species	4	16
Unknown protein/ best blast hit unknown protein	1	5
Transposable elements/ pseudogenes	4	27
MicroRNA of unknown function	1	1
Sum	10	49

The only directly analysed differentially expressed gene encodes a microRNA, miR-398a, that was shown to target copper/ zinc (Cu/ Zn) superoxide dismutases (CSD1 and CSD2) (Jagadeeswaran et al., 2009). Superoxide dismutases detoxify superoxide radicals and are therefore important for resistance to oxidative stress (Sunkar et al., 2006). It is possible, that *dpa4* plants somehow produce more superoxide radicals, and that miR398a is downregulated in *dpa4* to allow CSD expression, but there is no phenotypic evidence for this. Apart from the described phenotypes, *dpa4* plants look completely vital. Thus, no conclusions explaining the abnormal developmental phenotype of *dpa4* plants can be drawn from the downregulation of miR398a.

The same holds true for most of the sequence derived predicted functions of the remaining differentially expressed genes. Only two of them are related to genes that are involved in plant development (Tab. 10.2), *DEVIL 21* (*DVL21*) and *SHORT INTERNODES RELATED SEQUENCE 8* (*SRS8*).

DVL genes encode small polypeptides/ proteins of 40 to 144 amino acids containing a conserved domain named DVL or ROTUNDIFOLIA (RTF) domain. For DVL1-5 and DVL16 it has been shown that overexpression causes developmental phenotypic alterations, like altered silique shape (varying from broader tips of the siliques to broader bases), smaller rosette diameters with rounder leaves and smaller flowers. Loss-of-function mutants of *DVL1*, *DVL3* and *DVL16* display no phenotypic changes (Narita et al.,

2004; Wen et al., 2004). The traits affected in the *DVL* overexpressors largely overlap with the traits affected in *dpa4* and *35S::DPA4* respectively. Nevertheless, the trends observed for these traits are nearly opposite of expected if *DVL21* acts similar to other DVL domain containing proteins: *DVL21* was found to be upregulated in *dpa4*, which should then result in smaller flowers, but petals in *dpa4* are enlarged compared to wild type. Leaves in *dpa4* plants are also rounder as in *DVL* gene overexpressing plants, but leaf size is unaffected in *dpa4*. Abnormally shaped siliques are only found in *35S:DPA4*, where *DVL21* is expected to be downregulated. To test whether there is a relation between *DVL21* expression and the phenotype observed in *dpa4*, plants overexpressing *DVL21* will be constructed and *DVL21* expression levels will be analysed in *35S:DPA4* plants.

SHORT INTERNODES (*SHI*) genes built a small family of 10 genes in Arabidopsis, including the genes *SHI*, *STYLISH* (*STY1/2*) and *SRS3-8* (Kuusk et al., 2006). Kuusk et al. (2006) suggested *SRS8* to be a pseudogene, since the transcript was not detectable. The gene is localised to the pericentromeric region of chromosome 5 and the first part of the gene is identical to At1g35460, a basic helix-loop-helix (bHLH) family gene on another chromosome, pointing to a duplication event. Nevertheless, *SRS8* was found to be expressed in Col-0 at an average level in this study. In addition to that, expression was also observed in the AT-TAX project (Laubinger et al., 2008) as visualised by TileViz (www.weigelworld.org). Single mutants in *SHI* genes do not show abnormal phenotypes except for *sty1*, which displays abnormal gynoecium development including a broader replum and and wider styles. Double *sty1 sty2* mutants are also affected in leaf margin development and show enhance leaf serration, whereas leaves in *35S::STY1* plants are smooth, narrow and epinastic. Gynoecia in the double mutant show lobes developed from the style. Beside the epinastic leaves and the severe gynoecium development defect, *SHI* mutants and overexpressing phenotypes resemble the *DPA* related phenotype (Kuusk et al., 2002, 2006). However, no other *SHI* gene beside *SRS8* was found to be differentially regulated in *dpa4*. Further analysis will address the question whether *SRS8* is truly expressed and if the expression level is somehow related to the observed phenotypic alterations. For this purpose, *SRS8* expression will be analysed in *35S::DPA4* plants and a *35S::SRS8* line will be created. Since no T-DNA insertions are available in this locus, a reduction of *SRS8* might be achieved by RNA interference (RNAi).

Table 10.2: **Differentially expressed genes in *dpa4* with annotated function.** Differentially expressed genes were determined using RankProd at cut off $p < 0.05$. Functional annotations were derived from TAIR. The ✓sign in column 4 indicates that the RAV recognition motif CACCTG is present in the 3000 bp upstream of the gene. Fold change (FC) values are indicated as $\log_2$ values.

Gene	Functional prediction [1]	Confirmed function [2]	CACCTG	Up/Down [3]	FC
At2g26010	predicted PR protein	non	✓	Up	-1.80
At3g31902	pseudogene, MADS-box protein	non	✓	Up	-1.58
At2g04395	telomere maintenance	non	✓	Up	-1.38
At1g09995	Best Arabidopsis protein match is: helicase-related (At1g79890.1)	non		Up	-1.11
At3g23637	Contains DVL domain (DVL21), shoot development	non		Up	-1.22
At2g03445	MIR398A	Targets two Cu/Zn superoxide dismutases (CSD1 and CSD2) (Jagadeeswaran et al., 2009).		Up	-1.27
At1g61224	MIR842A, targets several Jacalin lectin family members.	non	✓	Down	1.58
At4g14811	MIR780A, targets CATION/H+ EXCHANGER 18 (CHX18) .	non	✓	Down	1.25
At4g09775	Best Arabidopsis protein match is: reverse transcriptase-related (At2g02650.1).	non		Down	1.28
At5g02360	Best Arabidopsis protein match is: DC1 domain-containing protein (At1g53340.1)	non		Down	1.29
At2g28755	Best Arabidopsis protein match is: NAD-dependent epimerase/dehydratase family protein (At2g28760.3)	non	✓	Down	1.50
At5g38378	Encodes a Plant thionin family protein.	non	✓	Down	1.26
At3g23680	Best Arabidopsis protein match is: F-box family protein (At3g23685.1)	non		Down	1.12
At5g33210	SHI-RELATED SEQUENCE 8 (SRS8), a member of SHORT INTERNODE (SHI) gene family. Arabidopsis thaliana has ten members that encode proteins with a RING finger-like zinc finger motif. SRS8 is a putative pseudogene (Kuusk et al., 2006).	non		Down	1.30
At4g02950	Best Arabidopsis protein match is: ubiquitin protein (At4g03360.1)	non	✓	Down	1.35
At3g01345	Contains InterPro domain glycoside hydrolase	non, downregulated in mutants of the SET domain protein, SDG4 (Cartagena et al., 2008).		Down	1.26
At5g39290	ARABIDOPSIS THALIANA EXPANSIN A26 (ATEXPA26)	non		Down	1.19
At5g39540	Best Arabidopsis protein match is: anac063 (Arabidopsis NAC (NAM/ATAF1, 2/CUC2) domain 63).	non		Down	1.40

[1] Function predicted by sequence similarity.
[2] Function confirmed by direct assay.
[3] Gene is up or down regulated in *dpa4*.

DPA4 was identified as a candidate for a putative role in development in the reverse genetics screen described in Part I. Phenotypes of *dpa4* T-DNA insertion lines and the overexpressor indicate that DPA4 is indeed involved in development, namely in the control of leaf margin shape and petal formation.

11.1 DPA4 controls leaf margin shape and petal size

Two independent T-DNA insertion lines with reduced *dpa4* expression display the same phenotype of stronger leaf serration and enlarged petals. The amount of remaining full-length transcript is correlated with the strength of the phenotype: leaves of *dpa4_1* plants, in which the full-length transcript is still detectable in low amounts, are less serrated than *dpa4_2* leaves. This observation can be considered as a strong indication that the reduction in *DPA4* expression is causing the described defects in leaf and petal development. Further evidence for the connection of *DPA4* to the observed serrated leaf margin phenotype is the absence of serrations in *35S::DPA4* plants.

11.1.1 Leaf margin defects in *dpa4*

Since overall leaf area and the number of cells per area are not altered in *dpa4* plants, the alterations in leaf margin shape can not simply be caused by general decrease in cell cycling, which has been shown to increase the depth of leaf serrations (Wang et al., 2000; Veylder et al., 2001). The changes in leaf margin shape seem to be rather caused by a change in growth direction regulation. This idea is further supported by the additional changes in overall leaf shape: *dpa4* leaf blades show weaker lateral expansion in the proximal and stronger lateral expansion at the distal part of the leaf compared to wild type.

11.1.2 Petal shape in *dpa4*

The role of *DPA4* in petal formation seems to be slightly different from the one in leaves, because petals are enlarged while cell numbers per area are stable as in leaves, indicating that enhanced cell proliferation takes place in petals. Despite of this, a similar alteration in shape is observed in petals as in leaves: petals are also more narrow in the proximal part and wider in the distal part compared to wild type, suggesting a role of *DPA4* in the direction of growth in petals as well.

11.1.3 Overexpression of *DPA4* results in pleiotropic developmental defects

Defects in *35S::DPA4* are more severe and affect additional phenotypic alterations compared to loss-of-function phenotypes. Three explanations for this observation are possible: either *DPA4* function is partially redundant (*e.g.* with At3g11580) or ectopic expression of *DPA4* leads to misexpression of targets in tissues where these are normally not affected by *DPA4*. The third possibility is that expression of *DPA4* at a high level causes the DPA4 protein to regulate genes which are usually not targets or targets of other RAV proteins. In the later case conclusions drawn from the overexpressor phenotype lead into wrong directions and do not facilitate understanding of DPA4 function. Further analysis of the crosses between At3g11580 T-DNA insertion lines and *dpa4* will reveal whether redundancies exist. The phenotype of *35S::DPA4* suggests auxin to be involved in the generation of the observed defects. The narrow shape of the leaves as well as their left handed twisting resemble the phenotype of the auxin influx-carrier *AUX1/LAX1-3* quadruple mutant (Bainbridge et al., 2008). These plants fail to establish auxin maxima to direct leaf initiation. In SD conditions, these plants also display defects in phyllotaxis (Bainbridge et al., 2008), which will also be investigated for *35S::DPA4* in future experiments.

A further hint to auxin signalling being impaired in *35S::DPA4* plants is the abnormal shape of siliques, which are similar to siliques of *NGA3* overexpressors. NGA proteins have been suggested to be involved in auxin signalling through elevation of *YUC2* and *YUC4* expression (Trigueros et al., 2009), which mediate auxin biosynthesis (Section

1.3.2) in the style. NGA proteins belong to the RAV family as well as DPA4, therefore this observation might support the idea of *DPA4* overexpression to cause targeting of non DPA4 targets. On the other hand, expression patterns in the carpel overlap between *DPA4* and *NGA1/2*, indicating that also redundancies are possible. For *NGA1* this overlap seems to be even more striking, since *NGA1* was found in the apex cluster described in Part I, indicating an overlapping expression in the shoot apex in general.

These two observed putative connections of DPA4 to auxin signalling are contradictory, since the leaf phenotype suggest DPA4 to negatively regulate auxin abundance while NGA proteins positively regulate auxin biosynthesis. It is impossible at the moment to conclude which scenario is more likely, but since local auxin concentration are in fact the shape determining factor in this signalling pathway, it might also be that two alleged contradictory functions exist in different tissues.

11.2 DPA acts in lateral organ boundary region, leaf sinus and floral primordia

Expression of *DPA4* in the boundary region between the shoot apex and the emerging organs suggests a role of *DPA4* early in development, for example in the establishment or maintenance of the lateral organ boundary. However, no tissue fusions or separation defects were observed in *dpa4*. Therefore, *DPA4* does not seem to be mainly involved in boundary formation but in the control and direction of organ outgrowth. A slight hint to this function is given by the fasciation of *35S::DPA4* stems, indicating that if DPA4 plays a role in tissue separation at all, this would be a negative one.

11.2.1 *DPA4* expression domains overlap with *CUC2* expression domains

Expression patterns of *DPA4* and *CUC2* are overlapping in almost all post embryonic parts of the plant. *CUC2*, as *DPA4*, is expressed in the boundary region of the SAM, in leaf sinuses, at the base of emerging floral organs and the septum region of carpels (Ishida et al., 2000; Nikovics et al., 2006). During embryo development *CUC2* is already expressed in the SAM (Aida et al., 1999), while no *DPA4* expression could be detected in

embryos in preliminary *in situ* hybridisation experiments (data not shown).

Alterations in *CUC2* transcript levels result in similar phenotypes as in *dpa4* plants, but in opposite direction: overexpression of a *miR164a* resistant version of *CUC2* results in enhanced leaf serration (Nikovics et al., 2006), whereas overexpression of *miR164a* causes smooth margins. Gymnoecium development is also affected in *cuc1/cuc2* double mutants (Ishida et al., 2000). Since these mutations are in L*er* background where siliques are broader than in Col-0, it is hard to compare phenotypes. Beside the developmental alterations of *CUC2* overexpression similar to *dpa4* plants, loss of CUC1/2 function results in tissue fusions in cotyledons, sepals and stamens (Aida et al., 1997).

Taken together, the data suggest that *DPA4* and *CUC2* might act together or in the same network controlling leaf margin formation. This interaction or mutual regulation would be limited to the unique functions of *CUC2* in organ outgrowth regulation (8.2.1), since phenotypic overlap was only observed for this characteristic. If the interaction is realised on transcript level (as one might expect for two transcription factors), *CUC2* expression should be altered in *dpa4* or vice versa. No elevation of *CUC2* expression was found in *dpa4* plants (data not shown), suggesting that, if a transcriptional regulation exists, *DPA4* might act downstream of *CUC2*. Measuring expression differences of *DPA4* in *cuc2* plants was rather difficult, since so far only L*er* alleles of *cuc2* were used and *DPA4* expression was found to be generally lower in this ecotype compared to Col-0 in preliminary qRT-PCR and *in situ* hybridisation experiments. In addition, there might still be some compensations by *CUC1*. Therefore, further experiments will focus on analysis in Col-0 background.

CUC2 is thought to act upstream of PIN1 mediated auxin transport (Section 8.2.1) to promote leaf serration formation. According to the proposed contradictory action, DPA4 should then have a negative effect of PIN1 mediated functions. This is supported by the *35S::DPA4* phenotype: smooth leaf margins and fasciated inflorescence stems were also observed in *pin1* mutant plants (Vernoux et al., 2000; Hay and Tsiantis, 2006).

11.3 Link to PcG mediated repression

DPA4 was included in the screen described in this study due to the presence of the PcG-repression associated histone mark H3K27me3 and could be shown to be upregulated

in *clf/swn* double mutants. From this result it is not possible to tell whether this upregulation is a direct cause of reduced H3K27me3 in *clf/swn* or an indirect cause due to upregulation of *DPA4* regulating genes. It is known that expression of the *KNOX* genes *STM* and *KNAT2* is regulated by the PcG proteins CLF/SWN and FIE (Katz et al., 2004; Schubert et al., 2006). Misexpression of *KNOX* genes is considered to cause the abnormal leaf phenotype in *clf* and *fie* mutant plants. Furthermore, the leaf shape controlling genes *miR164a*, *CUC2* and *PIN1* and the downstream of auxin signalling acting gene *ANT* are H3K27me3 targets. If *DPA4* is involved in some place in these networks, it is very likely that the upregulation in *clf/swn* is at least partially indirect.

Expression of *DPA4* is restricted to well defined areas in few tissues. The fact that PcG targets are on average more tissue specific than the rest of the genome suggest that PcG mediated regulation might be responsible for the restricted expression pattern of *DPA4*, as well as *CUC2* and partially also *miR164a* (Nikovics et al., 2006), which share this expression pattern. Preliminary in situ hybridisation experiments in *clf/swn* callus tissue indicate that *DPA4* expression is still restricted to discrete sub sets of cells in the callus (data not shown). If this result can be confirmed it would implicate that additional factors are responsible for the tissue specific expression of *DPA4*. Since this preservation of tissue specificity was observed for other H3K27me3 targets as well (Section 1.4), it might be considered as a further hint that H3K27me3 labelling is not the cause of tissue specificity but maybe a consequence. In this scenario, so far unknown factors might establish the tissue specific expression patterns observed for PcG targets and the labelling with H3K27me3 would take place afterwards to stabilise the expression pattern, maybe to prevent complete silencing of the targets genes as it is mediated by heterochromatin associated histone marks. This would be in accordance with the current observation that loss of H3K27me3 labelling is rather a consequence and not a cause for transcriptional activation (Adrian et al., 2010).

11.4 Putative *DPA4* target genes

The results of the expression array analysis for *dpa4* have to be verified by qRT-PCR, as analysis of such genome-wide experiments always includes false positive results. In particular, this analysis should include expression data of putative differentially expressed

genes in *35S::DPA4* plants to corroborate data from the loss-of-function mutant.

Leaf tissue was mostly excluded from the analysis, although adult leaves display the most obvious part of the abnormal phenotype in *dpa4* plants. Nevertheless, the proportion of affected tissue is lower in adult leaves based on qRT-PCR and *in situ* hybridisation data, which suggest a role of *DPA4* rather early in development. Young leafs and leaf primordia, where the proportion of affected tissue should be higher were included in apex enriched samples (Section 4.11).

So far no direct conclusions about the genetic network *DPA4* might be involved in can be drawn from the differentially expressed genes, since only one of them is characterised so far and is not involved in development (*miR398a*, involved in stress response). To conclude whether changes in expression of the two putative development involved genes *DVL21* and the SHI family gene *SRS8* can explain parts of the *dpa4* phenotype, further experiments are necessary (Section 10.6.2). Interestingly, other member of the SHI family (STY1 and STY2) act cooperatively with NGA proteins to regulate YUC mediated auxin synthesis (Trigueros et al., 2009). This fact supports the idea that DPA4 and NGA proteins might act redundantly or at least in very similar pathways.

If DPA4 is indeed a transcriptional repressor, it is not likely that the downregulated genes in *dpa4* are direct targets of DPA4. The fact that the AuxRE motif is slightly overrepresented in the promoters of these genes suggest that they might be indirectly regulated by DPA4 through ARF mediated auxin signalling (Section 1.3.2).

12 Conclusions and perspectives

In this study, *DPA4* was identified as an apex expressed H3K27me3 target regulating several aspects of Arabidopsis shoot development. Mainly affected traits are leaf and petal shape. Taking together the observed characteristics of *DPA4*, such as expression in organ boundary regions, similarities of *dpa4* plants to *miR164a* overexpressing plants and several hints to auxin signalling, lead to a current working model for DPA4 action: DPA4 is proposed to act downstream of CUC2 in organ boundaries (Fig. 12.1). Due to opposing effects of changes in *CUC2* and *DPA4* expression, *DPA4* expression should be repressed by CUC2. CUC2 was shown to be involved in the regulation of leaf margin auxin patterning that directs leaf serration formation (Kawamura et al., 2010). Kawamura et al. (2010) suggest that CUC2 is regulating *PIN1* expression or localisation and thus the formation of auxin maxima to direct outgrowth of serrations. Several indications suggest that DPA4 might act upstream of auxin signalling. *35S::DPA4* plants resemble *AUX1/LAX1-3* quadruple mutants impaired in auxin influx and putative indirect DPA4 targets display overrepresentation of AuxREs in their promoters. Possibly, DPA4 acts downstream of CUC2 and upstream of PIN1 or other auxin directing factors to repress leaf margin outgrowth through an inhibition of auxin maximum formation at the tip of the emerging tooth (Fig. 12.1).

To test this model, we are currently analysing the distribution of a GFP tagged version of PIN1 (Benková et al., 2003) in *dpa4* plants (in collaboration with Iris Leuz, MPIPZ, Cologne). Furthermore, auxin distributions will be monitored using the *DR5rev::GFP* (Benková et al., 2003) marker introduced to *35S::DPA4* and *dpa4* plants.

The putative regulation of *DPA4* by CUC2 will be tested by qRT-PCR measurements of *DPA4* expression in apex enriched *cuc2* and *mir164a* samples. *DPA4* expression will also be monitored by *in situ* hybridisation in these mutants, although it is challenging to perform truly quantitative analysis in such a restricted set of tissues.

Since several results indicate that DPA4 might act redundantly with other factors, the *dpa4_2* insertion will be introduced into loss-of-function mutants of other RAV proteins, in particular the closest relative, At3g11580.

To monitor DPA4 protein abundance, a translational GFP fusion for DPA4 under the

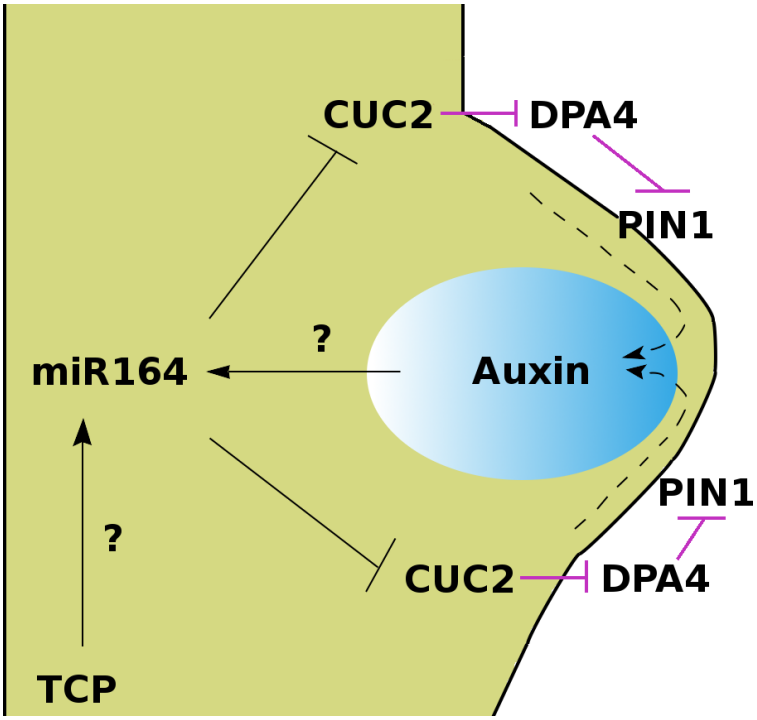


Figure 12.1: **Current working model for DPA4 action in leaf serration formation.** Leaf serrations are formed at auxin maxima, which are established by PIN1 mediated polar auxin transport. PIN1 expression and/or localisation was suggested to be regulated by CUC2, which is repressed by *miR164a*. *miR164a* is putatively regulated by TCP transcription factors and in a feedback loop by PIN1/auxin (Furutani et al., 2004). DPA4 might act downstream of CUC2 to repress PIN1 mediated establishing of auxin maxima. Interactions postulated during this study are depicted in pink. Figure modified after (Barkoulas et al., 2007)

control of its own promoter will be created. This will allow observation of cellular localisation of the protein to evaluate if DPA4 acts in the nucleus as expected for a transcription factor. Furthermore, the marker line will allow to distinguish expression patterns from protein abundance and thus clarify whether DPA4 protein acts only in organ boundaries or translocates into other domains.

Furthermore, it would be interesting to observe *DPA4* expression in later stages of leaf development. For this purpose, we are preparing a *DPA4* promoter reporter line driving expression of fluorescent marker proteins to test if *DPA4* expression persists at the leaf sinus and organ boundaries as the organs develop.

To further characterise the regulation of *DPA4* by PcG components, expression analysis by qRT-PCR and *in situ* hybridisation will be repeated. If the preliminary result that *DPA4* expression is still restricted to certain cells in *clf/swn* tissue can be confirmed,

further investigations will focus on the factors controlling this tissue specificity. This question is already addressed by Dr. Sara Farrona (MPIPZ, Cologne) for the control of *FT* expression and further studies on DPA4 will be performed in collaboration with her.

Adrian, J., Farrona, S., Reimer, J. J., Albani, M. C., Coupland, G., and Turck, F. (2010). cis-Regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis. *Plant Cell*, **22**(5), 1425–1440.

Adrian, J., Torti, S., and Turck, F. (2009). From decision to commitment: the molecular memory of flowering. *Mol Plant*, **2**(4), 628–642.

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *Plant Cell*, **9**(6), 841–857.

Aida, M., Ishida, T., and Tasaka, M. (1999). Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. *Development*, **126**(8), 1563–1570.

Al-Shahrour, F., Minguez, P., Tárraga, J., Medina, I., Alloza, E., Montaner, D., and Dopazo, J. (2007). FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res*, **35**(Web Server issue), W91–W96.

Al-Shahrour, F., Minguez, P., Tárraga, J., Montaner, D., Alloza, E., Vaquerizas, J. M., Conde, L., Blaschke, C., Vera, J., and Dopazo, J. (2006). BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res*, **34**(Web Server issue), W472–W476.

Bainbridge, K., Guyomarc'h, S., Bayer, E., Swarup, R., Bennett, M., Mandel, T., and Kuhlemeier, C. (2008). Auxin influx carriers stabilize phyllotactic patterning. *Genes Dev*, **22**(6), 810–823.

Barkoulas, M., Galinha, C., Grigg, S. P., and Tsiantis, M. (2007). From genes to shape: regulatory interactions in leaf development. *Curr Opin Plant Biol*, **10**(6), 660–666.

Barton, M. K. (2010). Twenty years on: the inner workings of the shoot apical meristem, a developmental dynamo. *Dev Biol*, **341**(1), 95–113.

Bengtsson, H., Simpson, K., Bullard, J., and Hansen, K. (2008). aroma.affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. Technical Report 745, Department of Statistics, University of California, Berkeley.

Benková, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertová, D., Jürgens, G., and Friml, J. (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell*, **115**(5), 591–602.

Bensmihen, S., To, A., Lambert, G., Kroj, T., Giraudat, J., and Parcy, F. (2004). Analysis of an activated ABI5 allele using a new selection method for transgenic Arabidopsis seeds. *FEBS Lett*, **561**(1-3), 127–131.

Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature*, **447**(7143), 407–412.

Blazquez, Green, Nilsson, Sussman, and Weigel (1998). Gibberellins promote flowering of arabidopsis by activating the LEAFY promoter. *Plant Cell*, **10**(5), 791–800.

Bleecker, A. B. and Patterson, S. E. (1997). Last exit: senescence, abscission, and meristem arrest in Arabidopsis. *Plant Cell*, **9**(7), 1169–1179.

Bomblies, K., Dagenais, N., and Weigel, D. (1999). Redundant enhancers mediate transcriptional repression of AGAMOUS by APETALA2. *Dev Biol*, **216**(1), 260–264.

Boss, P. K., Bastow, R. M., Mylne, J. S., and Dean, C. (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell*, **16 Suppl**, S18–S31.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N., and Coen, E. (1993). Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of Antirrhinum. *Cell*, **72**(1), 85–95.

Bratzel, F., López-Torrejón, G., Koch, M., Pozo, J. C. D., and Calonje, M. (2010). Keeping cell identity in arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Curr Biol*, **20**(20), 1853–1859.

Brioudes, F., Joly, C., Szécsi, J., Varaud, E., Leroux, J., Bellvert, F., Bertrand, C., and Bendahmane, M. (2009). Jasmonate controls late development stages of petal growth in Arabidopsis thaliana. *Plant J*, **60**(6), 1070–1080.

Byrne, M. E., Barley, R., Curtis, M., Arroyo, J. M., Dunham, M., Hudson, A., and Martienssen, R. A. (2000). Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis. *Nature*, **408**(6815), 967–971.

Byrne, M. E., Simorowski, J., and Martienssen, R. A. (2002). ASYMMETRIC LEAVES1 reveals knox gene redundancy in Arabidopsis. *Development*, **129**(8), 1957–1965.

Calonje, M., Sanchez, R., Chen, L., and Sung, Z. R. (2008). EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in Arabidopsis. *Plant Cell*, **20**(2), 277–291.

Candela, H., Martínez-Laborda, A., and Micol, J. L. (1999). Venation pattern formation in Arabidopsis thaliana vegetative leaves. *Dev Biol*, **205**(1), 205–216.

Cartagena, J. A., Matsunaga, S., Seki, M., Kurihara, D., Yokoyama, M., Shinozaki, K., Fujimoto, S., Azumi, Y., Uchiyama, S., and Fukui, K. (2008). The Arabidopsis SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. *Dev Biol*, **315**(2), 355–368.

Castillejo, C. and Pelaz, S. (2008). The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. *Curr Biol*, **18**(17), 1338–1343.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z. R., and Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development*, **131**(21), 5263–5276.

Cho, K. H., Jun, S. E., Lee, Y. K., Jeong, S. J., and Kim, G. T. (2007). Developmental Processes of Leaf Morphogenesis in Arabidopsis. *Plant Biology*, **50(3)**, 282–290.

Clark, S. E., Running, M. P., and Meyerowitz, E. M. (1995). CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development*, **121**(7), 2057–2067.

Clough, S. J. and Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J*, **16**(6), 735–743.

Coulibaly, I. and Page, G. P. (2008). Bioinformatic Tools for Inferring Functional Information from Plant Microarray Data II: Analysis Beyond Single Gene. *Int J Plant Genomics*, **2008**, 893941.

Dai, M., Wang, P., Boyd, A. D., Kostov, G., Athey, B., Jones, E. G., Bunney, W. E., Myers, R. M., Speed, T. P., Akil, H., Watson, S. J., and Meng, F. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Research*, **33**(20), e175.

Desprez, T., Juraniec, M., Crowell, E. F., Jouy, H., Pochylova, Z., Parcy, F., Höfte, H., Gonneau, M., and Vernhettes, S. (2007). Organization of cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, **104**(39), 15572–15577.

Disch, S., Anastasiou, E., Sharma, V. K., Laux, T., Fletcher, J. C., and Lenhard, M. (2006). The E3 ubiquitin ligase BIG BROTHER controls arabidopsis organ size in a dosage-dependent manner. *Curr Biol*, **16**(3), 272–279.

Dodsworth, S. (2009). A diverse and intricate signalling network regulates stem cell fate in the shoot apical meristem. *Dev Biol*, **336**(1), 1–9.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, **95**(25), 14863–14868.

Engelhorn, J. and Turck, F. (2010). Metaanalysis of ChIP-chip data. *Methods Mol Biol*, **631**, 185–207.

Farrona, S., Coupland, G., and Turck, F. (2008). The impact of chromatin regulation on the floral transition. *Semin Cell Dev Biol*, **19**(6), 560–573.

Fornara, F., de Montaigu, A., and Coupland, G. (2010). SnapShot: Control of flowering in Arabidopsis. *Cell*, **141**(3), 550, 550.e1–550, 550.e2.

Fraley, R. T., Rogers, S. G., Horsch, R. B., Eichholtz, D. A., Flick, J. S., Fink, C. L., Hoffmann, N. L., and Sanders, P. R. (1985). The SEV System: A New Disarmed Ti Plasmid Vector System for Plant Transformation. *Nat Biotech*, **3**(7), 629–635.

Furutani, M., Vernoux, T., Traas, J., Kato, T., Tasaka, M., and Aida, M. (2004). PIN-FORMED1 and PINOID regulate boundary formation and cotyledon development in Arabidopsis embryogenesis. *Development*, **131**(20), 5021–5030.

Gendall, A. R., Levy, Y. Y., Wilson, A., and Dean, C. (2001). The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell*, **107**(4), 525–535.

Göbel, U., Reimer, J., and Turck, F. (2010). Genome-wide mapping of protein-DNA interaction by chromatin immunoprecipitation and DNA microarray hybridization (ChIP-chip). Part B: ChIP-chip data analysis. *Methods Mol Biol*, **631**, 161–184.

Gómez-Mena, C., de Folter, S., Costa, M. M. R., Angenent, G. C., and Sablowski, R. (2005). Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. *Development*, **132**(3), 429–438.

Grennan, A. K. (2006). Variations on a theme. Regulation of flowering time in Arabidopsis. *Plant Physiol*, **140**(2), 399–400.

Grewal, S. I. S. and Elgin, S. C. R. (2007). Transcription and RNA interference in the formation of heterochromatin. *Nature*, **447**(7143), 399–406.

Grigg, S. P., Canales, C., Hay, A., and Tsiantis, M. (2005). SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. *Nature*, **437**(7061), 1022–1026.

Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A., and Gagliano, W. B. (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science*, **280**(5362), 446–450.

Hartigan, J. A. and Wong, M. A. (1979). A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**, 100–108.

Hay, A. and Tsiantis, M. (2006). The genetic basis for differences in leaf form between Arabidopsis thaliana and its wild relative Cardamine hirsuta. *Nat Genet*, **38**(8), 942–947.

Hepworth, S. R., Klenz, J. E., and Haughn, G. W. (2006). UFO in the Arabidopsis inflorescence apex is required for floral-meristem identity and bract suppression. *Planta*, **223**(4), 769–778.

Hicke, L., Schubert, H. L., and Hill, C. P. (2005). Ubiquitin-binding domains. *Nat Rev Mol Cell Biol*, **6**(8), 610–621.

Hong, F., Breitling, R., McEntee, C. W., Wittner, B. S., Nemhauser, J. L., and Chory, J. (2006). RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics*, **22**(22), 2825–2827.

Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, **409**(6819), 525–529.

Hu, Y., Xie, Q., and Chua, N.-H. (2003). The Arabidopsis auxin-inducible gene ARGOS controls lateral organ size. *Plant Cell*, **15**(9), 1951–1961.

Hu, Y. X., Wang, Y. X., Liu, X. F., and Li, J. Y. (2004). Arabidopsis RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. *Cell Res*, **14**(1), 8–15.

Huala, E. and Sussex, I. M. (1992). LEAFY Interacts with Floral Homeotic Genes to Regulate Arabidopsis Floral Development. *Plant Cell*, **4**(8), 901–913.

Hugouvieux, V., Barber, C. E., and Daniels, M. J. (1998). Entry of Xanthomonas campestris pv. campestris into hydathodes of Arabidopsis thaliana leaves: a system for studying early infection events in bacterial pathogenesis. *Mol Plant Microbe Interact*, **11**(6), 537–543.

Ikeda, M. and Ohme-Takagi, M. (2009). A novel group of transcriptional repressors in Arabidopsis. *Plant Cell Physiol*, **50**(5), 970–975.

Irish, V. F. (2008). The Arabidopsis petal: a model for plant organogenesis. *Trends Plant Sci*, **13**(8), 430–436.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**(2), 249–264.

Ishida, T., Aida, M., Takada, S., and Tasaka, M. (2000). Involvement of CUP-SHAPED COTYLEDON genes in gynoecium and ovule development in Arabidopsis thaliana. *Plant Cell Physiol*, **41**(1), 60–67.

Jagadeeswaran, G., Saini, A., and Sunkar, R. (2009). Biotic and abiotic stress down-regulate miR398 expression in Arabidopsis. *Planta*, **229**(4), 1009–1014.

Jang, S., Torti, S., and Coupland, G. (2009). Genetic and spatial interactions between FT, TSF and SVP during the early stages of floral induction in Arabidopsis. *Plant J*, **60**(4), 614–625.

Jenuwein, T. and Allis, C. D. (2001). Translating the histone code. *Science*, **293**(5532), 1074–1080.

Jöcker, A., Hoffmann, F., Groscurth, A., and Schoof, H. (2008). Protein function prediction and annotation in an integrated environment powered by web services (AFAWE). *Bioinformatics*, **24**(20), 2393–2394.

Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res*, **27**(2), 470–478.

Katz, A., Oliva, M., Mosquna, A., Hakim, O., and Ohad, N. (2004). FIE and CURLY LEAF polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development. *Plant J*, **37**(5), 707–719.

Kawamura, E., Horiguchi, G., and Tsukaya, H. (2010). Mechanisms of leaf tooth formation in Arabidopsis. *Plant J*, **62**(3), 429–441.

Kayes, J. M. and Clark, S. E. (1998). CLAVATA2, a regulator of meristem and organ development in Arabidopsis. *Development*, **125**(19), 3843–3851.

Khatri, P. and Drăghici, S. (2005). Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, **21**(18), 3587–3595.

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., and Gruissem, W. (2003). Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO J*, **22**(18), 4804–4814.

Kotake, T., Takada, S., Nakahigashi, K., Ohto, M., and Goto, K. (2003). Arabidopsis TERMINAL FLOWER 2 gene encodes a heterochromatin protein 1 homolog and represses both FLOWERING LOCUS T to regulate flowering time and several floral homeotic genes. *Plant Cell Physiol*, **44**(6), 555–564.

Koyama, T., Furutani, M., Tasaka, M., and Ohme-Takagi, M. (2007). TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in Arabidopsis. *Plant Cell*, **19**(2), 473–484.

Kuhlemeier, C. (2007). Phyllotaxis. *Trends Plant Sci*, **12**(4), 143–150.

Kunst, L., Klenz, J. E., Martinez-Zapater, J., and Haughn, G. W. (1989). AP2 Gene Determines the Identity of Perianth Organs in Flowers of Arabidopsis thaliana. *Plant Cell*, **1**(12), 1195–1208.

Kuusk, S., Sohlberg, J. J., Eklund, D. M., and Sundberg, E. (2006). Functionally redundant SHI family genes regulate Arabidopsis gynoecium development in a dose-dependent manner. *Plant J*, **47**(1), 99–111.

Kuusk, S., Sohlberg, J. J., Long, J. A., Fridborg, I., and Sundberg, E. (2002). STY1 and STY2 promote the formation of apical tissues during Arabidopsis gynoecium development. *Development*, **129**(20), 4707–4717.

Kuzmichev, A., Margueron, R., Vaquero, A., Preissner, T. S., Scher, M., Kirmizis, A., Ouyang, X., Brockdorff, N., Abate-Shen, C., Farnham, P., and Reinberg, D. (2005). Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. *Proc Natl Acad Sci U S A*, **102**(6), 1859–1864.

Laubinger, S., Marchal, V., Gourrierec, J. L., Gentilhomme, J., Wenkel, S., Adrian, J., Jang, S., Kulajta, C., Braun, H., Coupland, G., and Hoecker, U. (2006). Arabidopsis SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. *Development*, **133**(16), 3213–3222.

Laubinger, S., Zeller, G., Henz, S. R., Sachsenberg, T., Widmer, C. K., Naouar, N., Vuylsteke, M., Schölkopf, B., Rätsch, G., and Weigel, D. (2008). At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis thaliana. *Genome Biol*, **9**(7), R112.

Laufs, P., Jonak, C., and Traas, J. (1998). Cells and domains: Two views of the shoot meristem in Arabidopsis. *Plant Physiology and Biochemistry*, **36**(1-2), 33–45.

Laux, T., Mayer, K. F., Berger, J., and Jürgens, G. (1996). The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. *Development*, **122**(1), 87–96.

Lawson, E. J. and Poethig, R. S. (1995). Shoot development in plants: time for a change. *Trends Genet*, **11**(7), 263–268.

Lee, D.-K., Geisler, M., and Springer, P. S. (2009). LATERAL ORGAN FUSION1 and LATERAL ORGAN FUSION2 function in lateral organ separation and axillary meristem formation in Arabidopsis. *Development*, **136**(14), 2423–2432.

Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R., and Dean, C. (2002). Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science*, **297**(5579), 243–246.

Leyser, O. (2005). The fall and rise of apical dominance. *Curr Opin Genet Dev*, **15**(4), 468–471.

Leyser, O. (2006). Dynamic integration of auxin transport and signalling. *Curr Biol*, **16**(11), R424–R433.

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C. A., Ito, T., Meyerowitz, E., and Yu, H. (2008a). A repressor complex governs the integration of flowering signals in Arabidopsis. *Dev Cell*, **15**(1), 110–120.

Li, J., Hansen, B. G., Ober, J. A., Kliebenstein, D. J., and Halkier, B. A. (2008b). Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. *Plant Physiol*, **148**(3), 1721–1733.

Li, Q., Wrange, O., and Eriksson, P. (1997). The role of chromatin in transcriptional regulation. *Int J Biochem Cell Biol*, **29**(5), 731–742.

Li, Y., Zheng, L., Corke, F., Smith, C., and Bevan, M. W. (2008c). Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. *Genes Dev*, **22**(10), 1331–1336.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**(6648), 251–260.

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., Grossniklaus, U., and Köhler, C. (2006). Different Polycomb group complexes regulate common target genes in Arabidopsis. *EMBO Rep*, **7**(9), 947–952.

Malgieri, A., Kantzari, E., Patrizi, M. P., and Gambardella, S. (2010). Bone marrow and umbilical cord blood human mesenchymal stem cells: state of the art. *Int J Clin Exp Med*, **3**(4), 248–269.

Mandel, M. A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M. F. (1992). Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. *Nature*, **360**(6401), 273–277.

Martin-Tryon, E. L., Kreps, J. A., and Harmer, S. L. (2007). GIGANTEA acts in blue light signaling and has biochemically separable roles in circadian clock and flowering time regulation. *Plant Physiol*, **143**(1), 473–486.

Mathieu, J., Yant, L. J., Mürdter, F., Küttner, F., and Schmid, M. (2009). Repression of flowering by the miR172 target SMZ. *PLoS Biol*, **7**(7), e1000148.

Matser, V., Davies, B., and Waites, R. (2010). The role of class I TCP genes in determining leaf shape and size. In *21ST INTERNATIONAL CONFERENCE ON ARABIDOPSIS RESEARCH*.

McGhee, J. D. and Felsenfeld, G. (1980). Nucleosome structure. *Annu Rev Biochem*, **49**, 1115–1156.

Meijer, A. H., van Dijk, E. L., and Hoge, J. H. (1996). Novel members of a family of AT hook-containing DNA-binding proteins from rice are identified through their in vitro

interaction with consensus target sites of plant and animal homeodomain proteins. *Plant Mol Biol*, **31**(3), 607–618.

Michniewicz, M., Zago, M. K., Abas, L., Weijers, D., Schweighofer, A., Meskiene, I., Heisler, M. G., Ohno, C., Zhang, J., Huang, F., Schwab, R., Weigel, D., Meyerowitz, E. M., Luschnig, C., Offringa, R., and Friml, J. (2007). Antagonistic regulation of PIN phosphorylation by PP2A and PINOID directs auxin flux. *Cell*, **130**(6), 1044–1056.

Miwa, H., Kinoshita, A., Fukuda, H., and Sawa, S. (2009). Plant meristems: CLAVATA3/ESR -related signaling in the shoot apical meristem and the root apical meristem. *Journal of Plant Research*, **122**, 31–39. 10.1007/s10265-008-0207-3.

Mizukami, Y. and Fischer, R. L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proc Natl Acad Sci U S A*, **97**(2), 942–947.

Mönke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.-P., Bäumlein, H., and Conrad, U. (2004). Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta*, **219**(1), 158–166.

Morey, L. and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci*, **35**(6), 323–332.

Nag, A., King, S., and Jack, T. (2009). miR319a targeting of TCP4 is critical for petal growth and development in Arabidopsis. *Proc Natl Acad Sci U S A*, **106**(52), 22534–22539.

Nakai, T., Kato, K., Shinmyo, A., and Sekine, M. (2006). Arabidopsis KRPs have distinct inhibitory activity toward cyclin D2-associated kinases, including plant-specific B-type cyclin-dependent kinase. *FEBS Lett*, **580**(1), 336–340.

Narita, N. N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., Goodrich, J., and Tsukaya, H. (2004). Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in Arabidopsis thaliana. *Plant J*, **38**(4), 699–713.

Nath, U., Crawford, B. C. W., Carpenter, R., and Coen, E. (2003). Genetic control of surface curvature. *Science*, **299**(5611), 1404–1407.

Ng, M. and Yanofsky, M. F. (2001). Activation of the Arabidopsis B class homeotic genes by APETALA1. *Plant Cell*, **13**(4), 739–753.

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. *Plant Cell*, **18**(11), 2929–2945.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J. J., Goldberg, R. B., and Fischer, R. L. (1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell*, **11**(3), 407–416.

Ori, N., Eshed, Y., Chuck, G., Bowman, J. L., and Hake, S. (2000). Mechanisms that control knox gene expression in the Arabidopsis shoot. *Development*, **127**(24), 5523–5532.

Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. *Nature*, **425**(6955), 257–263.

Parcy, F., Nilsson, O., Busch, M. A., Lee, I., and Weigel, D. (1998). A genetic framework for floral patterning. *Nature*, **395**(6702), 561–566.

Petricka, J. J., Clay, N. K., and Nelson, T. M. (2008). Vein patterning screens and the defectively organized tributaries mutants in Arabidopsis thaliana. *Plant J*, **56**(2), 251–263.

Pien, S. and Grossniklaus, U. (2007). Polycomb group and trithorax group proteins in Arabidopsis. *Biochim Biophys Acta*, **1769**(5-6), 375–382.

Pilot, G., Stransky, H., Bushey, D. F., Pratelli, R., Ludewig, U., Wingate, V. P. M., and Frommer, W. B. (2004). Overexpression of GLUTAMINE DUMPER1 leads to hypersecretion of glutamine from Hydathodes of Arabidopsis leaves. *Plant Cell*, **16**(7), 1827–1840.

Prigge, M. J. and Wagner, D. R. (2001). The arabidopsis serrate gene encodes a zinc-finger protein required for normal shoot development. *Plant Cell*, **13**(6), 1263–1279.

Rehrauer, H., Aquino, C., Gruissem, W., Henz, S. R., Hilson, P., Laubinger, S., Naouar, N., Patrignani, A., Rombauts, S., Shu, H., de Peer, Y. V., Vuylsteke, M., Weigel, D., Zeller, G., and Hennig, L. (2010). AGRONOMICS1: a new resource for Arabidopsis transcriptome profiling. *Plant Physiol*, **152**(2), 487–499.

Reimer, J. J. and Turck, F. (2010). Genome-wide mapping of protein-DNA interaction by chromatin immunoprecipitation and DNA microarray hybridization (ChIP-chip). Part A: ChIP-chip molecular methods. *Methods Mol Biol*, **631**, 139–160.

Reinhardt, D., Pesce, E.-R., Stieger, P., Mandel, T., Baltensperger, K., Bennett, M., Traas, J., Friml, J., and Kuhlemeier, C. (2003). Regulation of phyllotaxis by polar auxin transport. *Nature*, **426**(6964), 255–260.

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B. K., and Yu, G. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**(5499), 2105–2110.

Ringrose, L. and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet*, **38**, 413–443.

Rivals, I., Personnaz, L., Taing, L., and Potier, M.-C. (2007). Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, **23**(4), 401–407.

Romanel, E. A. C., Schrago, C. G., nago, R. M. C., Russo, C. A. M., and Alves-Ferreira, M. (2009). Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PLoS One*, **4**(6), e5791.

Rosso, M. G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol Biol*, **53**(1-2), 247–259.

Sanchez-Pulido, L., Devos, D., Sung, Z. R., and Calonje, M. (2008). RAWUL: a new ubiquitin-like domain in PRC1 ring finger proteins that unveils putative plant and worm PRC1 orthologs. *BMC Genomics*, **9**, 308.

Sawa, M., Nusinow, D. A., Kay, S. A., and Imaizumi, T. (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. *Science*, **318**(5848), 261–265.

Scarpella, E., Barkoulas, M., and Tsiantis, M. (2010). Control of leaf and vein development by auxin. *Cold Spring Harb Perspect Biol*, **2**(1), a001511.

Scarpella, E., Marcos, D., Friml, J., and Berleth, T. (2006). Control of leaf vascular patterning by polar auxin transport. *Genes Dev*, **20**(8), 1015–1027.

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J. U. (2005). A gene expression map of Arabidopsis thaliana development. *Nat Genet*, **37**(5), 501–506.

Schoof, H., Lenhard, M., Haecker, A., Mayer, K. F., Jürgens, G., and Laux, T. (2000). The stem cell population of Arabidopsis shoot meristems in maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell*, **100**(6), 635–644.

Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., and Scott, R. J. (2006). The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development*, **133**(2), 251–261.

Schubert, D., Clarenz, O., and Goodrich, J. (2005). Epigenetic control of plant development by Polycomb-group proteins. *Curr Opin Plant Biol*, **8**(5), 553–561.

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J*, **25**(19), 4638–4649.

Schug, J., Schuller, W.-P., Kappen, C., Salbaum, J. M., Bucan, M., and Stoeckert, C. J. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol*, **6**(4), R33.

Schultz, E. A. and Haughn, G. W. (1991). LEAFY, a Homeotic Gene That Regulates Inflorescence Development in Arabidopsis. *Plant Cell*, **3**(8), 771–781.

Schwartz, Y. B. and Pirrotta, V. (2008). Polycomb complexes and epigenetic states. *Curr Opin Cell Biol*, **20**(3), 266–273.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J. D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitzel, T., Katagiri, F., Glazebrook, J., Law, M., and Goff, S. A. (2002). A high-throughput Arabidopsis reverse genetics system. *Plant Cell*, **14**(12), 2985–2994.

Simon, J. A. and Kingston, R. E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol*, **10**(10), 697–708.

Simpson, G. G. (2004). The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of Arabidopsis flowering time. *Curr Opin Plant Biol*, **7**(5), 570–574.

Stepanova, A. N., Robertson-Hoyt, J., Yun, J., Benavente, L. M., Xie, D.-Y., Dolezal, K., Schlereth, A., Jürgens, G., and Alonso, J. M. (2008). TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell*, **133**(1), 177–191.

Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, **403**(6765), 41–45.

Sturn, A. (2001). Cluster Analysis for Large Scale Gene Expression Studies. Master's thesis, Graz University of Technology and The Institute for Genomic Research Maryland.

Sturn, A., Quackenbush, J., and Trajanoski, Z. (2002). Genesis: cluster analysis of microarray data. *Bioinformatics*, **18**(1), 207–208.

Sung, S., He, Y., Eshoo, T. W., Tamada, Y., Johnson, L., Nakahigashi, K., Goto, K., Jacobsen, S. E., and Amasino, R. M. (2006). Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1. *Nat Genet*, **38**(6), 706–710.

Sunkar, R., Kapoor, A., and Zhu, J.-K. (2006). Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell*, **18**(8), 2051–2065.

Suzuki, M., Wang, H. H.-Y., and McCarty, D. R. (2007). Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol*, **143**(2), 902–911.

Swaminathan, K., Peterson, K., and Jack, T. (2008). The plant B3 superfamily. *Trends Plant Sci*, **13**(12), 647–655.

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature*, **462**(7274), 799–802.

Szécsi, J., Joly, C., Bordji, K., Varaud, E., Cock, J. M., Dumas, C., and Bendahmane, M. (2006). BIGPETALp, a bHLH transcription factor is involved in the control of Arabidopsis petal size. *EMBO J*, **25**(16), 3912–3920.

Takada, S., Hibara, K., Ishida, T., and Tasaka, M. (2001). The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation. *Development*, **128**(7), 1127–1135.

Telfer, A., Bollman, K., and Poethig, R. (1997). Phase change and the regulation of trichome distribution in Arabidopsis thaliana. *Development*, **124**(3), 645–654.

Thiriet, C. and Hayes, J. J. (2006). Histone dynamics during transcription: exchange of H2A/H2B dimers and H3/H4 tetramers during pol II elongation. *Results Probl Cell Differ*, **41**, 77–90.

Trigueros, M., Navarrete-Gómez, M., Sato, S., Christensen, S. K., Pelaz, S., Weigel, D., Yanofsky, M. F., and Ferrándiz, C. (2009). The NGATHA genes direct style development in the Arabidopsis gynoecium. *Plant Cell*, **21**(5), 1394–1409.

Tsukaya, H. (2006). Mechanism of leaf-shape determination. *Annu Rev Plant Biol*, **57**, 477–496.

Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. *Annu Rev Plant Biol*, **59**, 573–594.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.-L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R. A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet*, **3**(6), e86.

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. (1999). Activation and repression of transcription by auxin-response factors. *Proc Natl Acad Sci U S A*, **96**(10), 5844–5849.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., and Coupland, G. (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science*, **303**(5660), 1003–1006.

Vernoux, T., Besnard, F., and Traas, J. (2010). Auxin at the shoot apical meristem. *Cold Spring Harb Perspect Biol*, **2**(4), a001487.

Vernoux, T., Kronenberger, J., Grandjean, O., Laufs, P., and Traas, J. (2000). PIN-FORMED 1 regulates cell fate at the periphery of the shoot apical meristem. *Development*, **127**(23), 5157–5165.

Veylder, L. D., Beeckman, T., Beemster, G. T., Krols, L., Terras, F., Landrieu, I., van der Schueren, E., Maes, S., Naudts, M., and InzÃ©, D. (2001). Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis. *Plant Cell*, **13**(7), 1653–1668.

Vroemen, C. W., Mordhorst, A. P., Albrecht, C., Kwaaitaal, M. A. C. J., and de Vries, S. C. (2003). The CUP-SHAPED COTYLEDON3 gene is required for boundary and shoot meristem formation in Arabidopsis. *Plant Cell*, **15**(7), 1563–1577.

Wang, G., Zhang, Z., Angenent, G. C., and Fiers, M. (2010). New aspects of CLAVATA2, a versatile gene in the regulation of Arabidopsis development. *J Plant Physiol*.

Wang, H., Zhou, Y., Gilmer, S., Whitwill, S., and Fowke, L. C. (2000). Expression of the plant cyclin-dependent kinase inhibitor ICK1 affects cell division, plant growth and morphology. *Plant J*, **24**(5), 613–623.

Wang, J.-W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. *Cell*, **138**(4), 738–749.

Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F., and Meyerowitz, E. M. (1992). LEAFY controls floral meristem identity in Arabidopsis. *Cell*, **69**(5), 843–859.

Wen, J., Lease, K. A., and Walker, J. C. (2004). DVL, a novel class of small polypeptides: overexpression alters Arabidopsis development. *Plant J*, **37**(5), 668–677.

Widom, J. (1989). Toward a unified model of chromatin folding. *Annu Rev Biophys Biophys Chem*, **18**, 365–395.

Wilson, R. N., Heckman, J. W., and Somerville, C. R. (1992). Gibberellin Is Required for Flowering in Arabidopsis thaliana under Short Days. *Plant Physiol*, **100**(1), 403–408.

Wisman, E., Cardon, G. H., Fransz, P., and Saedler, H. (1998). The behaviour of the autonomous maize transposable element En/Spm in Arabidopsis thaliana allows efficient mutagenesis. *Plant Mol Biol*, **37**(6), 989–999.

Woody, S. T., Austin-Phillips, S., Amasino, R. M., and Krysan, P. J. (2007). The WiscDsLox T-DNA collection: an arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. *J Plant Res*, **120**(1), 157–165.

Xu, B., Li, Z., Zhu, Y., Wang, H., Ma, H., Dong, A., and Huang, H. (2008). Arabidopsis genes AS1, AS2, and JAG negatively regulate boundary-specifying genes to promote sepal and petal development. *Plant Physiol*, **146**(2), 566–575.

Xu, L. and Shen, W.-H. (2008). Polycomb silencing of KNOX genes confines shoot stem cell niches in Arabidopsis. *Curr Biol*, **18**(24), 1966–1971.

Yang, L., Liu, Z., Lu, F., Dong, A., and Huang, H. (2006). SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J*, **47**(6), 841–850.

Yant, L., Mathieu, J., Dinh, T. T., Ott, F., Lanz, C., Wollmann, H., Chen, X., and Schmid, M. (2010). Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. *Plant Cell*, **22**(7), 2156–2170.

Yi, Y. and Jack, T. (1998). An intragenic suppressor of the Arabidopsis floral organ identity mutant apetala3-1 functions by suppressing defects in splicing. *Plant Cell*, **10**(9), 1465–1477.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V., Pellegrini, M., Goodrich, J., and Jacobsen, S. E. (2007a). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLoS Biol*, **5**(5), e129.

Zhang, X., Germann, S., Blus, B. J., Khorasanizadeh, S., Gaudin, V., and Jacobsen, S. E. (2007b). The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat Struct Mol Biol*, **14**(9), 869–871.

Zimmermann, R. and Werr, W. (2005). Pattern formation in the monocot embryo as revealed by NAM and CUC3 orthologues from Zea mays L. *Plant Mol Biol*, **58**(5), 669–685.

Table A.1: **Sample list of "Developmental series" expression data set. List was obtained from www.weigelworld.org.**

DEVELOPMENTAL SERIES (TRIPLICATE ARRAYS, COLUMBIA BACKGROUND)

RNAs provided by MPI Tübingen (Schmid, Lohmann; development baseline), Univ. Leicester (Twell; pollen), MPI Cologne (Weisshaar; seed & silique development), Univ. Pennsylvania (Poethig; phase change), CAGE consortium (Kuiper; common reference samples) and Univ. Utrecht (Scheres; development on MS agar). Probes prepared and hybridized by Markus Schmid, Jan Lohmann and Monika Demar at the MPI Tübingen (Dept. Weigel).

#	Sample ID	Experiment description	Genotype	Tissue	Age	Photoperiod	Substrate
1	**ATGE_1**	development baseline	Wt	cotyledons	7 days	continuous light	soil
2	**ATGE_2**	development baseline	Wt	hypocotyl	7 days	continuous light	soil
3	**ATGE_3**	development baseline	Wt	roots	7 days	continuous light	soil
4	**ATGE_4**	development baseline	Wt	shoot apex, vegetative + young leaves	7 days	continuous light	soil
5	**ATGE_5**	development baseline	Wt	leaves 1 +2	7 days	continuous light	soil
6	**ATGE_6**	development baseline	Wt	shoot apex, vegetative	7 days	continuous light	soil
7	**ATGE_7**	development baseline	Wt	seedling, green parts	7 days	continuous light	soil
8	**ATGE_8**	development baseline	Wt	shoot apex, transition (before bolting)	14 days	continuous light	soil
9	**ATGE_9**	development baseline	Wt	roots	17 days	continuous light	soil
10	**ATGE_10**	development baseline	Wt	rosette leaf #4, 1 cm long	10 days	continuous light	soil
11	**ATGE_11**	development baseline	*gl1*-T	rosette leaf #4, 1 cm long	10 days	continuous light	soil
12	**ATGE_12**	development baseline	Wt	rosette leaf #2	17 days	continuous light	soil
13	**ATGE_13**	development baseline	Wt	rosette leaf #4	17 days	continuous light	soil
14	**ATGE_14**	development baseline	Wt	rosette leaf #6	17 days	continuous light	soil
15	**ATGE_15**	development baseline	Wt	rosette leaf #8	17 days	continuous light	soil
16	**ATGE_16**	development baseline	Wt	rosette leaf #10	17 days	continuous light	soil
17	**ATGE_17**	development baseline	Wt	rosette leaf #12	17 days	continuous light	soil
18	**ATGE_18**	development baseline	*gl1*-T	rosette leaf #12	17 days	continuous light	soil
19	**ATGE_19**	development baseline	Wt	leaf 7, petiole	17 days	continuous light	soil
20	**ATGE_20**	development baseline	Wt	leaf 7, proximal half	17 days	continuous light	soil
21	**ATGE_21**	development baseline	Wt	leaf 7, distal half	17 days	continuous light	soil
22	**ATGE_22**	development baseline	Wt	developmental drift, entire rosette after transition to flowering, but before bolting	21 days	continuous light	soil
23	**ATGE_23**	development baseline	Wt	as above	22 days	continuous light	soil
24	**ATGE_24**	development baseline	Wt	as above	23 days	continuous light	soil
25	**ATGE_25**	development baseline	Wt	senescing leaves	35 days	continuous light	soil
26	**ATGE_26**	development baseline	Wt	cauline leaves	21+days	continuous light	soil
27	**ATGE_27**	development baseline	Wt	stem, 2nd internode	21+days	continuous light	soil
28	**ATGE_28**	development baseline	Wt	1st node	21+days	continuous light	soil
29	**ATGE_29**	development baseline	Wt	shoot apex, inflorescence (after bolting)	21 days	continuous light	soil
30	**ATGE_31**	development baseline	Wt	flowers stage 9	21+days	continuous light	soil
31	**ATGE_32**	development baseline	Wt	flowers stage 10/11	21+days	continuous light	soil
32	**ATGE_33**	development baseline	Wt	flowers stage 12	21+days	continuous light	soil
33	**ATGE_34**	development baseline	Wt	flowers stage 12, sepals	21+days	continuous light	soil
34	**ATGE_35**	development baseline	Wt	flowers stage 12, petals	21+days	continuous light	soil
35	**ATGE_36**	development baseline	Wt	flowers stage 12, stamens	21+days	continuous light	soil
36	**ATGE_37**	development baseline	Wt	flowers stage 12, carpels	21+days	continuous light	soil
37	**ATGE_39**	development baseline	Wt	flowers stage 15	21+days	continuous light	soil
38	**ATGE_40**	development baseline	Wt	flowers stage 15, pedicels	21+days	continuous light	soil

#	Sample ID	Experiment description	Genotype	Tissue	Age	Photoperiod	Substrate
39	**ATGE_41**	development baseline	Wt	flowers stage 15, sepals	21+days	continuous light	soil
40	**ATGE_42**	development baseline	Wt	flowers stage 15, petals	21+days	continuous light	soil
41	**ATGE_43**	development baseline	Wt	flowers stage 15, stamen	21+days	continuous light	soil
42	**ATGE_45**	development baseline	Wt	flowers stage 15, carpels	21+days	continuous light	soil
43	**ATGE_46**	development baseline	*clv3-7*	shoot apex, inflorescence (after bolting)	21+days	continuous light	soil
44	**ATGE_47**	development baseline	*lfy-12*	shoot apex, inflorescence (after bolting)	21+days	continuous light	soil
45	**ATGE_48**	development baseline	*ap1-15*	shoot apex, inflorescence (after bolting)	21+days	continuous light	soil
46	**ATGE_49**	development baseline	*ap2-6*	shoot apex, inflorescence (after bolting)	21+days	continuous light	soil
47	**ATGE_50**	development baseline	*ap3-6*	shoot apex, inflorescence (after bolting)	21+days	continuous light	soil
48	**ATGE_51**	development baseline	*ag-12*	shoot apex, inflorescence (after bolting)	21+days	continuous light	soil
49	**ATGE_52**	development baseline	*ufo-1*	shoot apex, inflorescence (after bolting)	21+days	continuous light	soil
50	**ATGE_53**	development baseline	*clv3-7*	flower stage 12; multi-carpel gynoeceum; enlarged meristem; increased organ number	21+days	continuous light	soil
51	**ATGE_54**	development baseline	*lfy-12*	flower stage 12; shoot characteristics; most organs leaf-like	21+days	continuous light	soil
52	**ATGE_55**	development baseline	*ap1-15*	flower stage 12; sepals replaced by leaf-like organs, petals mostly lacking, 2° flowers	21+days	continuous light	soil
53	**ATGE_56**	development baseline	*ap2-6*	flower stage 12; no sepals or petals	21+days	continuous light	soil
54	**ATGE_57**	development baseline	*ap3-6*	flower stage 12; no petals or stamens	21+days	continuous light	soil
55	**ATGE_58**	development baseline	*ag-12*	flower stage 12; no stamens or carpels	21+days	continuous light	soil
56	**ATGE_59**	development baseline	*ufo-1*	flower stage 12; filamentous organs in whorls two and three	21+days	continuous light	soil
57	**ATGE_73**	pollen	Wt	mature pollen	6 wk	continuous light	soil
58	**ATGE_76**	seed & silique development	Wt	siliques, w/ seeds stage 3; mid globular to early heart embryos	8 wk	long day (16/8)	soil
59	**ATGE_77**	seed & silique development	Wt	siliques, w/ seeds stage 4; early to late heart embryos	8 wk	long day (16/8)	soil
60	**ATGE_78**	seed & silique development	Wt	siliques, w/ seeds stage 5; late heart to mid torpedo embryos	8 wk	long day (16/8)	soil
61	**ATGE_79**	seed & silique development	Wt	seeds, stage 6, w/o siliques; mid to late torpedo embryos	8 wk	long day (16/8)	soil
62	**ATGE_81**	seed & silique development	Wt	seeds, stage 7, w/o siliques; late torpedo to early walking-stick embryos	8 wk	long day (16/8)	soil
63	**ATGE_82**	seed & silique development	Wt	seeds, stage 8, w/o siliques; walking-stick to early curled cotyledons embryos	8 wk	long day (16/8)	soil
64	**ATGE_83**	seed & silique development	Wt	seeds, stage 9, w/o siliques; curled cotyledons to early green cotyledons embryos	8 wk	long day (16/8)	soil
65	**ATGE_84**	seed & silique development	Wt	seeds, stage 10, w/o siliques; green cotyledons embryos	8 wk	long day (16/8)	soil
66	**ATGE_87**	phase change	Wt	vegetative rosette	7 days	short day (10/14)	soil
67	**ATGE_89**	phase change	Wt	vegetative rosette	14 days	short day (10/14)	soil
68	**ATGE_90**	phase change	Wt	vegetative rosette	21 days	short day (10/14)	soil
69	**ATGE_91**	comparison with CAGE	Wt	leaf	15 days	long day (16/8)	1x MS agar, 1% sucrose
70	**ATGE_92**	comparison with CAGE	Wt	flower	28 days	long day (16/8)	soil
71	**ATGE_93**	comparison with CAGE	Wt	root	15 days	long day (16/8)	1x MS agar, 1% sucrose
72	**ATGE_94**	development on MS agar	Wt	root	8 days	continuous light	1x MS agar
73	**ATGE_95**	development on MS agar	Wt	root	8 days	continuous light	1x MS agar, 1% sucrose
74	**ATGE_96**	development on MS agar	Wt	seedling, green parts	8 days	continuous light	1x MS agar
75	**ATGE_97**	development on MS agar	Wt	seedling, green parts	8 days	continuous light	1x MS agar, 1% sucrose
76	**ATGE_98**	development on MS agar	Wt	root	21 days	continuous light	1x MS agar
77	**ATGE_99**	development on MS agar	Wt	root	21 days	continuous light	1x MS agar, 1% sucrose
78	**ATGE_100**	development on MS agar	Wt	seedling, green parts	21 days	continuous light	1x MS agar
79	**ATGE_101**	development on MS agar	Wt	seedling, green parts	21 days	continuous light	1x MS agar, 1% sucrose

Table A.2: **Assignment of "Developmental series" samples to tissue categories as used in this study. For cluster analysis samples were ordered according to their appearance in this list.**

Tissue group	Sample
Root	ATGE99 ATGE98 ATGE93 ATGE94 ATGE95 ATGE3 ATGE9
Stem	ATGE28 ATGE27 ATGE2
Leaf and whole plants	ATGE101 ATGE97 ATGE100 ATGE96 ATGE7 ATGE1 ATGE87 ATGE90 ATGE89 ATGE91 ATGE5 ATGE19 ATGE20 ATGE21 ATGE16 ATGE18 ATGE17 ATGE12 ATGE13 ATGE11 ATGE10 ATGE14 ATGE15 ATGE23 ATGE24 ATGE22 ATGE26 ATGE25
Apex	ATGE6 ATGE4 ATGE8 ATGE29 ATGE52 ATGE48 ATGE47 ATGE46 ATGE51 ATGE50 ATGE49
Flowers and floral organs	ATGE92 ATGE32 ATGE33 ATGE37 ATGE35 ATGE34 ATGE36 ATGE39 ATGE45 ATGE40 ATGE42 ATGE41 ATGE43 ATGE31 ATGE59 ATGE53 ATGE57 ATGE56 ATGE58 ATGE55 ATGE54
Seeds	ATGE76 ATGE77 ATGE78 ATGE73 ATGE84 ATGE79 ATGE81 ATGE82 ATGE83

Table A.3: **T-DNA insertion lines used in this study.** Internal names were given to the lines according to ordering date to ease handeling of lines with short names.

Gene	Insertion line	NASC ID	Internal code	Location in gene	LP-Primer	RP-Primer	Product LP/RP [bp]	Product LB/RP [bp]
At1g69690	SALK_011491	N511491	D1	Exon	AGAACCACGTAAGCCCATCTC	TCAAATGAACTCCACTACCGC	1132	600-900
	WiscDsLox442E5	N511491	G9	Exon	CGTATATTTTGATTCTTTGAACAGC	ATGATGATCGAAGCTTTGGTG	1126	596-896
At2g16210	SALK_012566	N512566	E1	Exon	GCTGGCAATCTCGACAATATC	CCCAAAGTTCTAAAAGGACCG	1166	537-837
	SALK_123864	N623864	F1	Exon	AGAATAAGAGGCGCGGATTAG	TGTATCGAAGGATCAAGTGCC	1058	448-748
At1g26780	SALK_025235	N525235	J1	Promoter	AGATTCATTTCGGACTCGTCC	CTTCAGCTGGTCTCCAATGTC	1126	494-794
At1g75710	SALK_048268	N548268	C2	Exon	CCGCCTTTAAGAAGGTCAAAC	CGTTAGATCTCGTTGAGCCAG	1027	465-765
	SALK_144673	N644673	D2	Exon	CTTGCTTGCATTCATGGAATC	CGTCTCTAAAGCTGCAAATGG	1064	468-768
	SALK_027882	N527882	B5	Exon	CCGCCTTTAAGAAGGTCAAAC	CGTTAGATCTCGTTGAGCCAG	1027	512-812
	SALK_060915	N560915	C10	Intron	ATCCGTTCGATCTGACAGATG	TCTCCGGATGTTACGAATGTC	1178	576-876
	SAIL_1222_C07	N844739	D10	Intron	CTCAACGAGATCTAACGCCAG	CTCCGAAACTGCAATTAAACG	1160	605-905
	SAIL_81_D10	N803840	E10	Intron	GAGAAGTTTTCCCCAAGCTTG	AGATTTTCCTTGGAAACCGTG	1253	605-905
	SALK_042064	N542064	F10	Intron	CAGCAAATCATCAGCTCTTCC	GATTTTCGTAAATGCGAATGG	1129	566-866
	SALK_092650	N592650	G10	Promoter	CAAACTTCAAGTGTAACTCCACC	AGGTGTTGAAATCGAAGATGG	1082	480-780
	SALK_139626	N639626	H10	Promoter	CAAACTTCAAGTGTAACTCCACC	AGGTGTTGAAATCGAAGATGG	1082	480-780
	SALK_046220C	N546220	B12	Intron	GATTTTCGTAAATGCGAATGG	CAGCAAATCATCAGCTCTTCC	1129	566-866
At5g46590	SALK_078798	N578798	F2	Exon	GTTGCAAATGGGAAGTTCATG	ACCGAATAAACCGGTTTGATC	1122	475-775
	SALK_078797C	N663058	G2	Exon	AACAACAAATCCAGCATCGAG	ACCGAATAAACCGGTTTGATC	1265	595-895
At5g24580	SALK_132335	N632335	B3	Exon	ATCCCTTCCAAAACATTCCAC	GCTTGTGGATTACGTCTACCG	1128	610-910
At1g52410	SALK_151393	N651393	E3	Exon	AACCAGAAGACACGTCATTGC	CTGATGCTTCTGAGGCTTACG	992	474-774
At1g24020	SALK_022306C	N654546	F3	Exon	TTCTTCATGTGGAGGTTGAGG	ATGGCAAGTCTCAATTGATCG	1152	478-778
At5g51590	SALK_124619C	N656945	I3	Exon	CTGTGTGTGTACATTGGCCTC	CTGACCCGCTATAAAACTCCC	1161	552-852
	SALK_049264C	N662411	J3	Exon	GCTCGCGTGCTATTTGTATTC	ATTGGGGAGTTTGATTGAAGG	1129	563-863
	SALK_024656	N524656	A5	Exon	TGAAATTAAGATTCCGCCATG	CCAAAGCCTACAGATTCCTCG	1113	479-779
At1g13290	SALK_148869C	N667588	A4	Exon	TTCCCACAATTTTTCTCATGC	CATGGGCTTGTCACTCGTAAC	1062	438-738
At1g78170	SAIL_197_H05	N809279	D4	Exon	ATCTGCTCATTTCCTCACACG	CGTGTGTACGTGTTATTCCCC	1079	519-819
At1g54020	SALK_047900C	N660415	C5	Exon	ACTTCCTCAAAACTTTTGGCC	TAACGAAAGCTTGTTGAGCAG	972	449-749
	SALK_104816C	N660542	D5	Exon	TTCTAATGAGTGCGTATGGCC	TTTCTGCTCAACAAGCTTTCG	1063	498-798
At3g55110	SALK_003029	N503029	G5	Promoter	GAAATCTTTCAGGAGTTCGCC	CGAATCACGATTGTCAAATGTC	1091	569-869
	SALK_100187	N600187	H5	Exon	AGAAGAGACCCCAAGCTAACG	TCACAGAGTTCGCACTTGATG	1110	555-855
	SALK_003029	N503029	J11	Promoter	GAAATCTTTCAGGAGTTCGCC	CGAATCACGATTGTCAAATGTC	1091	569-869
At1g62540	SALK_098896	N598896	I5	Exon	TTTGCCACTGCTGGATTTTAG	TGGTTTCAAGGAATGTGAAGC	1127	574-874
	SALK_080561C	N678718	A12	Intron	ATAGCCATGGTGTCAGACCAC	TGCGTGTCTTTATAACACGTTTC	1202	569-869
At1g19830	SALK_054428	N554428	D6	Exon	GTGGTGAAATTGGAGCAACTC	GTCTACGTAGGAGGGAATCGG	1130	504-804
	SALK_054420	N554420	E6	Exon	ACCCCAAAACCATAACTCAGC	TATCCTGGAATCAGCATTTCG	1163	572-872
At4g27590	SALK_044439	N544439	B7	Exon	CTTCCTCGTCGATCACTTGAG	AACTCGATGTGTTAGCGATGG	984	440-740
At5g57130	SALK_083141	N583141	C7	Exon	TCTATTTCGTGTTTATGGCGG	CTTGAAACTCCTTGCTGATCG	1130	469-769
At2g32280	SALK_088331	N588331	H7	Exon	CCAAATCGGAGATTTGCTTTC	AGCCTCATGTGCTTCACCTAC	1206	600-900
At1g70560	SALK_067278	N567278	D8	Exon	GGTGAAGGACAAGGAGGTAGC	AGTTTCTGCACATCATGGAGG	1082	516-816
	SALK_022743C	N665532	E8	Exon	TCTTTTCCATTACAACGTGGG	CGTCAAGACCATGACATCATG	1088	536-836
At1g62500	SALK_022164	N522164	G8	Exon	TGTGCATTTTTGAATTTGAATAGTG	CATTCCTCTTGCACTTCAAGC	1056	491-791

Gene	Insertion line	NASC ID	Internal code	Location in gene	LP-Primer	RP-Primer	Product LP/RP [bp]	Product LB/RP [bp]
	SAIL_3_D11	CS800202	H8	Exon	GTTACAACTCCTCCGGGACTC	AAGCTTGAAGCTCCTTTTATCG	1282	605-905
At2g16580	SALK_072739	N572739	I8	Exon	GGAGAATTCATGCATTCGTTG	GAGGGGTCTTCTCGAAATGTC	1119	525-825
	SALK_056098	N669732	J8	Promoter	ACTCGTGTTACGAGTCGGATG	TATTGGATCTCTGTGGGAACG	1015	443-743
At1g60060	SALK_026005	N526005	A2	Promoter	TTGCAAATTTGGTATTCTTGG	TTTTCCCCAAAACAAACTCAC	1149	574-874
	SALK_132906	N632906	B2	Promoter	TCAATTTGGCAGTTTTGTCAG	AGGCAAGATCCTCCAGAAAAC	1161	507-807
	SAIL_851_B08	CS838012	H9	Exon	TGTCCAGAATTTTAGTCCAATGC	CAACGCTTCTCCTTCTGACAC	984	449-749
At1g31310	SAIL_98_E04	N804714	H4	Exon	TGCCGCAAACTTACTCACTTC	AAGGAGTGTGTCAGTGATCGC	1153	555-855
	GK_481D08	N446124	I9	Exon	AAACGCTTCCTTCTTCTACCG	TTATGTACTCCGATCATCACGC	1088	440-740
At3g28220	GK-116E03	N411091	B10	Exon	TATCGAACGCACAAAAGATCC	TAACGTCCATTTGTAAAGGCG	1226	601-901
At1g65370	SALK_104078C	N657269	E5	Exon	CCATTGATGGCCTAAAGAGTG	AACCATGTGAGTTTGGTGCTC	1035	441-741
	SALK_111532	N611532	F5	Exon	AACAGTCAACACGTTGAACCC	CTGACGGTGAAGTCTTATCGC	1081	473-773
	SALK_092544	N592544	J10	Exon	ACCAATTCTTTCCTCCCATTG	TCGAAGATGTCGAAAATGTCC	1070	528-828
	SALK_044365	N544365	A11	Promoter	AACGAGGGAGAGAAGACGAAG	TGCAACCCTAGAGAGCAAAAC	1084	459-759
At1g77200	SALK_123815	N623815	J2	Promoter	TAGCTCGAGGAAGACAACTGC	CGAGACCTAAGAAAACATCGC	1101	459-759
	SALK_096307C	N657742	A3	Promoter	TAGCTCGAGGAAGACAACTGC	CGAGACCTAAGAAAACATCGC	1101	462-762
At1g02190	SAIL_1142F07	N841962	E4	Promoter	ACTTGAAGCTACCAAGGGGAC	CGAGCTGTTCTGTTCTGGTTC	1049	505-805
	SALK_132241	N632241	F4	Promoter	GTAGAGAAACTCGACGGGACC	TCCAAAATGCTGATTTCGTTC	1100	553-853
At1g03710	SALK_080449	N660975	I4	Promoter	CATTCCACACTGACACACTGG	TCATGTTTGGCTTGAAGAAGG	1246	589-889
	SALK_083639	N583639	J4	Exon	AATGTAAGCGGCGTAGACATG	GGCTAAGGAATCTCCTCATGG	1027	466-766
At1g73620	SALK_100586C	N661048	J5	Promoter	TTCAGAGAGCCTTCAAACACC	GCTTGGGAGACAAGAAGAAGG	1138	480-780
	SALK_149186	N649186	A6	Intron	TCATAGGCTGGTTTGATTCATG	TTTCAACTTATTTCTTGTACAAACCG	1065	541-841
At2g39330	SALK_114673C	N659211	F6	Promoter	TAACGAACTTTGTACCGGTCG	GAATGATAACATGCAGCCAGG	1139	502-802
	SALK_109697	N609697	G6	Exon	CCGACTCCTATAATTCCAGCC	ATTTCGGTGAAATTGTGCAAC	1104	589-889
At3g12460	SALK_013751	N513751	E7	Promoter	GTCGGTAGTCGCTGTCTCATC	GTACTGGTGCTCGCTATCACC	1032	457-757
	SALK_093000	N593000	F7	Promoter	CCTCAACAATCTTCTCAAGCG	ATCAAATGTTTTTGCACCCAG	1009	434-734
At5g44620	SALK_057031	N557031	G7	Promoter	CACATAGCTTCCTCAGCATCC	TTTCTTTTGGTGATGAGTCCG	1253	606-906
At5g51850	SALK_131680C	N660187	I7	Promoter	AAGCCTGCAAATTTCACAATG	CCCAAAGAAAACAAAAGGACC	1052	459-759
At4g02670	SALK_062011	N562011	J7	Promoter	TCGTCTTTAGGGCGAAATAGC	TGCAAAGACTGTTTGATTCCC	1017	430-730
At3g13662	SALK_052848	N552848	B8	Promoter	TCGAATGTTAGTGGATCGTCC	AAGTGCAAACGATTCCACAAC	1110	466-766
	SAIL_748_E04	CS833424	C8	Exon	AAAATTTTGAATCACAAAATATGACG	AACCTCATCCTTATCTTCGCC	1108	449-749
At5g19730	SALK_016457	N516457	F8	Promoter	TCGCCGTTGCAATATTTTATC	ATGTTGTTAGATGCGGTCGAG	1005	455-755
At5g49330	SALK_104164C	N65513	G3	Intron	TCAGACCAATGGTGAAGGTTC	TGACGTCAGAAAAATTGCATG	1167	518-818
	SALK_024386C	N655370	H3	Intron	TTGTTCTTGGCTCCAGAAGAC	TAATCGTCAAGTTGCATTCCC	1103	536-836
	GK-291D01	N427877	J9	Exon	ACAATGAGTGGCAAAGTGGTC	ATGTACAATCCACTTGCTCGG	1216	549-849
At1g24070	SALK_009744	N509744	D7	Intron	AACATCGATGGTCTTGTGGTC	TATGAGAAAAGCAACCGCTTG	1179	507-807
At5g06250	SALK_150707	N650707	D3	Exon	TCTTCTGGTTGTCGTGTGTTG	AGGACACGAACAATTCATTCG	954	438-738
	SALK_088181C	N666657	I10	Intron	ACTAACAAAGTGGCCATGTGC	TCTTATCATCAAATCCACGCC	1053	501-801
	SALK_045786	N545786	I11	Intron	TCCTAGTCTCCTCTGCTGCTG	GTCTCGAGAAACAGCGTCATC	1222	599-899
At5g10420	SALK_045655	N545655	C4	Intron	CGTTTGCCACTCGAACTCTAC	TTTTCAGCTGCATGATCACTG	1068	454-754
	SALK_012094	N512094	E11	Exon	TCTAGCGGATGATTTTTGGTG	TTTCCTCTCAACAAATGTCGG	1069	461-761
	SALK_138885	N638885	F11	Intron	GAGATGATGATTCCACTTGCC	CCTCATTGTCCCAGTCACATC	1014	464-764
At4g25020	SALK_091758C	N65917	B6	Exon	GGCAGCACCTGTAACCTCTAG	ACAGGAAAGCTCTCGAAATCC	1127	530-830
	SALK_076904	N576904	C6	Exon	GGCAGCACCTGTAACCTCTAG	ACAGGAAAGCTCTCGAAATCC	1127	593-893

Gene	Insertion line	NASC ID	Internal code	Location in gene	LP-Primer	RP-Primer	Product LP/RP [bp]	Product LB/RP [bp]
At1g03720	SAIL_13_B08.v1	CS800599	H6	Exon	CTTAAGTTAGACGGGGCCATC	TGTTAGTTTGATTATTTAAATGGTTGTG	1051	471-771
	SALK_060104	N560105	I6	Promoter	TTTGCTTCCATCTCCAAATTC	TCTTTTCTGTGCGATTTGTTG	1129	524-824
At1g51460	SALK_025689	N52689	J6	Exon	TCTCTTAGGGGTTCATGCATG	TGAAAACAGGTGTTTTCTCGC	1173	544-844
	SALK_046735C	N662346	A7	Exon	GGCAAAGTACTTGTCAATGGC	CCCCTTTTTCCTTTCTGTGAC	1155	484-784
At5g17030	SALK_114099C	N667076	A8	Promoter	TTCACGAAGGATGGATTTGAG	TCCGTAAGGATGCACTTGAAC	1164	579-879
At5g20740	SALK_068607	N568607	A9	Promoter	GGGTGGATTAGGAAATGAAGC	GGTTCAGCTAAAGAGTTCCGG	1066	513-813
At1g71050	SALK_048115C	N665962	B9	Exon	TGGAGCCTTGAAAAATGAATG	AGATCAGATCACAATCACCGC	1127	593-893
At1g28290	SAIL_890C04	CS840070	C9	Promoter	ATCACGAATCAGATGCCTGAC	GAAGCACCAAAGTGCTACGAG	1104	559-859
At2g28790	WiscDsLox289	CS850431	D9	Exon	CTGCAAGCCTCGAATAAAAAC	AAATTCATTGCCTAACCACCC	1235	548-848
At5g18460	SALK_115026	N615026	H2	Exon	GGATTGATACCAGTCCAATTTC	CACTCCTCCAACCCGAACTA	1037	-
	SALK_134225	N634225	I2	Exon	CCAAGCATTAACGACTTGAGG	GTGAAGTCGTCAACACACGTG	1151	474-774
	SALK_134225C	N673523	E9	Exon	CCAAGCATTAACGACTTGAGG	GTGAAGTCGTCAACACACGTG	1151	474-774
	WisLox507F07	N859342	F9	Exon	CCCTATCAGAATTTTCATGCAC	CGCGGTTGACTAATACGACTAC	1220	610-910
	SALK_144533	N644533	G11	Intron	AGGCGACTATAAACGTTTGGG	AAGAGGAGATATGGCTGCTCC	1128	592-892
	SALK_061393	N561393	H11	Intron	AGGCGACTATAAACGTTTGGG	AAGAGGAGATATGGCTGCTCC	1128	487-787
At2g01610	SAIL_1158_B02	N842715	A13	Exon	TAACATGAGAACGACCCATCC	CGTTTTCTTTGTTCTTGTCCG	1229	572-872
At4g25320	SAIL_404_B09	N818706	-	Promoter	TTTCTCGGCCTACCTCTCTTC	GAGTTTGTGAAAGTTTGGGAGG	1097	554-854
	SAIL_663_B02	N828857	-	Promoter	AGCTTCTTCCATGGAGGAATG	TGTTGGTTGATTTTATTTGGAAG	1092	572-872
At3g11580	SAIL_878_A05	N839490	-	Exon	AGAAACTTGCGTTTGTGTTGC	AAGCTGGAAAGCAAAACTGAC	1097	485-785
	WiscDsLox387C07	N853929	-	Exon	AGAAACTTGCGTTTGTGTTGC	AAGCTGGAAAGCAAAACTGAC	1097	569-869

Table A.4: **Primers for amplification of coding regions to create overexpressors.** To forward primers attB1 extension (5´-GGGGACAAGTTTGTACAAAAAAGCAGGCTCC-3´) was added, to reverse primers attB2 tail (5´-GGGGACCACTTTGTACAAGAAAGCTGGGTC-3´) was added

Gene	forward primer	reverse primer
At1g69690	ATGGATCCGGATCCGGATCATAACC	CTAGGAATGATGACTGGTGCTTCCA
At1g75710	ATGGCTTTACTAACTTTCTTG	TCAAGATTCTAAAACTTTGTAAATA
At5g46590	ATGTTCTTGTCACACAAAGAATACA	CTAGGAGAAATCTGAGTAACCG
At1g24020	ATGGGGTTGAGTGGTGTT	TTAGGCACTAGTTTGCTT
At5g51590	ATGGAGGAGAGAGAAGGAACTAACA	TCAGCTTGGAACCTCGGTGTCA
At1g78170	ATGGTTTCTCGACAAGAA	CTAGCATAGAAGCCTAAG
At1g62540	ATGGCACCAGCTCAAAAC	TTAGAGGATATGGGAAGG
At1g19830	ATGGCGATTATAAACCGGAGC	CTAAATATATGTTGGCTGAAGCATA
At2g16580	ATGTCCATTCTCAAGAAA	TCATCGGATCATGGAAGT
At1g31310	ATGGCTGACCAAAGTGGT	CTAATTATTATGGCGAGA
At1g65370	ATGTATGTGGAATGTCTCTCAT	TTAAAACATTTTCCTAGA
At1g77200	ATGACCGAGTCATCCATT	CTAAGGAAAAAGGGGGCC
At1g02190	ATGGCGTCGAGGCCCGGA	TCATAGAGGAGATGGTGGGAGAG
At1g73620	ATGACCAATGAGAAATGTGAAGTTG	TCAACGGCCGCGGTGAGG
At5g19730	ATGCCCAAACTCAATTCA	TTAGAGTTTGATCCATTC
At5g06250	ATGTCAGTCAACCATTACTCCACAG	TTATAAAAGAGTTAAAATTACCATG
At5g20740	ATGGCTCCTACACAAAAT	TCAAAGATGTACGTCGTG
At1g71050	ATGGGAGCTCTTGATTCT	TTACATAACGGTGCAAGC
At5g18460	ATGGAGGTGAATGTTCTTGCATCC	TCAATTACACCTAGGATTAAACCCC
At4g16447	ATGAGTTCAATAGCAAGAGACCGAA	TTAGAGATGCAGAGACGTGTCAAC
At3g15680	ATGAGCAGACCCGGAGAT	TTAGAAAGAGGTTCTGTT

Table A.5: **Primers for amplification cDNA fragments via qRT-PCR.** Start positions indicated are positions of first base pair amplified in mature mRNA.

Gene	forward primer	reverse primer	product lenght [bp]	start position [bp]
DPA1	GGTGGTACGTTTGACTCGGT	GGGAATTATGATGCAGGTTGA	1364	231
DPA2	ACGTGCTTCAGGCAGGCACA	AGGATTAAACCCCGGCCCACCA	1385	235
DPA3	GCTCGACCCGCGACATTCGT	TCGTCTCTCACCGTGGCCGT	136	242
DPA4	ACCGCCACATGGCTCCACGA	ACGGCGGCTCCATAGTGGGA	527	29
DPA5	TGGTGGTAGTGGCGGCGGTA	TCTCCGCCACCGGTCGGTAA	939	176
DPA7	AATGCGGCAGAGGAAGGCGG	TTGCCCCGAACGCTTGACCG	670	265
DPA8	TCGATCACCATCCTGGCTAGCG	ACCTTGAAAGAGCCTAGCATAGAAGCC	551	230
DPA9	TGAAACTCCTCTTCCTCCTCCTGGT	AGCGAGTGAGCCGTCAGGGT	247	220
DPA10	TGCGCCTTGGTTTCCCGTCC	TGGACTCATGGTGGCGCTGTG	1001	231
DPA11	ACGAGTCACCGACGCTTGGC	ACCGAAGCGGTCTTTCGCCG	240	214
DPA12	GCATCGCTTGATGCGGTAGGC	ACACCCACCCCCAAAGTAAAAACACAA	1181	312
DPA13	AGGTCCAACTATCACACCGGTGA	GACCCCGGTCGTTTGCCTCG	126	237
CLV2	CAGGGGCGTTGGCCGGAAAT	ACTACGAGCACGAGCTGAGCA	1229	122

Table A.6: **Primers for amplification of templates for in-situ probe synthesis.** For anti-sense probes T7 polymerase promoter sequence was added to reverse primer, for sense probes, the sequence was added to the forward primer. Start positions indicated are positions of first base pair amplified in mature mRNA.

Probe name	forward primer	reverse primer	probe lenght [bp]	start position [bp]
DPA4 2	CGGGAAACTCAACCGCCTCGTC	ACGGCGGCTCCATAGTGGGA	459	231
DPA4 3	ATTGGCTGGCGCAGACGTGG	TGGCCGTAGTAACCGTCGGG	335	542

Table A.7: **Genes in PcG Apex Cluster.** Descriptions were downloaded form "The Arabidopsis Information Resource" (www.arabidopsis.org), genes which were considered as candidates in the screen are marked with a ✓icon

Locus Identifier	Annotation	Candidate?
At1g02190	CER1 protein	✓
At1g03710	cysteine protease inhibitor	✓
At1g03720	cathepsin-related	✓
	zinc finger (C2H2 type) family protein	✓
At1g19830	auxin-responsive protein	✓
At1g24020	MLP423 (MLP-LIKE PROTEIN 423)	✓
At1g24070	ATCSLA10 (Cellulose synthase-like A10); transferase	✓
At1g26780	MYB117 (myb domain protein 117); transcription factor	✓
At1g28290	AGP31 (ARABINOGALACTAN-PROTEIN 31); structural constituent of cell wall	✓
At1g31310	hydroxyproline-rich glycoprotein family protein	✓
At1g51460	ABC transporter family protein	✓
At1g52410	TSA1 (TSK-ASSOCIATING PROTEIN 1); calcium ion binding / protein binding	✓
At1g54020	myrosinase-associated protein	✓
At1g60060	similar to unknown protein [Arabidopsis thaliana] (TAIR:At5g53900.2); similar to unnamed protein product [Vitis vinifera] (GB:CAO40921.1); contains domain PTHR13902:SF3 (PTHR13902:SF3); contains domain PTHR13902 (PTHR13902)	✓
At1g62500	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	✓
At1g62540	flavin-containing monooxygenase family protein / FMO family protein	✓
At1g65370	meprin and TRAF homology domain-containing protein / MATH domain-containing protein	✓
At1g69690	TCP family transcription factor	✓
At1g70560	alliinase C-terminal domain-containing protein	✓
At1g71050	heavy-metal-associated domain-containing protein / copper chaperone (CCH)-related	✓
At1g73620	thaumatin-like protein	✓
At1g75710	zinc finger (C2H2 type) family protein	✓
At1g77200	AP2 domain-containing transcription factor TINY	✓
At1g78170	similar to unknown protein [Arabidopsis thaliana] (TAIR:At1g22250.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO61724.1)	✓
At2g01610	invertase/pectin methylesterase inhibitor family protein	✓
At2g16210	transcriptional factor B3 family protein	✓
At2g16580	auxin-responsive protein	✓
At2g28790	osmotin-like protein	✓
At2g32280	similar to unknown protein [Arabidopsis thaliana] (TAIR:At4g21310.1); similar to unknown [Populus trichocarpa] (GB:ABK92874.1); contains InterPro domain Protein of unknown function DUF1218 (InterPro:IPR009606)	✓
At2g39330	jacalin lectin family protein	✓
At3g12460	3′-5′ exonuclease/ nucleic acid binding	✓
At3g13662	disease resistance-responsive protein-related / dirigent protein-related	✓
At3g15680	zinc finger (Ran-binding) family protein	✓
At3g28220	meprin and TRAF homology domain-containing protein / MATH domain-containing protein	✓
At3g55110	ABC transporter family protein	✓
At4g02670	ATIDD12 (ARABIDOPSIS THALIANA INDETERMINATE(ID)-DOMAIN 12); nucleic acid binding / transcription factor/ zinc ion binding	✓
At4g16447	unknown protein	✓
At4g25020	KOW domain-containing protein / D111/G-patch domain-containing protein	✓
At4g27590	copper-binding protein-related	✓
At5g06250	transcription factor	✓
At5g10420	antiporter	✓
At5g17030	UDP-glucoronosyl/UDP-glucosyl transferase family protein	✓
At5g18460	similar to unknown protein [Arabidopsis thaliana] (TAIR:At1g23340.1); similar to unknown protein [Arabidopsis thaliana] (TAIR:At1g23340.2); similar to unnamed protein product [Vitis vinifera] (GB:CAO38766.1); contains InterPro domain Protein of unknown function DUF239	✓
At5g19730	pectinesterase family protein	✓
At5g20740	invertase/pectin methylesterase inhibitor family protein	✓
At5g24580	copper-binding family protein	✓
At5g44620	CYP706A3 (cytochrome P450	✓
At5g46590	ANAC096 (Arabidopsis NAC domain containing protein 96); transcription factor	✓
At5g49330	AtMYB111 (myb domain protein 111); DNA binding / transcription factor	✓
At5g51590	DNA-binding protein-related	✓
At5g51850	similar to unknown protein [Arabidopsis thaliana] (TAIR:At4g25430.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO17701.1)	✓
At5g57130	protein binding	✓
At1g06080	ADS1 (DELTA 9 DESATURASE 1); oxidoreductase	

Locus Identifier	Annotation	Candidate?
At1g11600	CYP77B1 (cytochrome P450)	
At1g13710	CYP78A5 (cytochrome P450)	
At1g17200	integral membrane family protein	
At1g23380	KNAT6 (Knotted-like Arabidopsis thaliana 6); DNA binding / transcription factor	
At1g26310	CAL (CAULIFLOWER); DNA binding / transcription factor	
At1g32240	KAN2 (KANADI 2); DNA binding / transcription factor	
At1g49430	LACS2 (LONG-CHAIN ACYL-COA SYNTHETASE 2)	
At1g52400	BGL1 (BETA-GLUCOSIDASE HOMOLOG 1); hydrolase	
At1g53160	SPL4 (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 4); DNA binding / transcription factor	
At1g57820	ORTH2/VIM1 (VARIANT IN METHYLATION 1); DNA binding / chromatin binding / double-stranded methylated DNA binding / histone binding / methyl-CpG binding / methyl-CpNpG binding / methyl-CpNpN binding	
At1g65620	AS2 (ASYMMETRIC LEAVES 2)	
At1g68480	JAG (JAGGED); nucleic acid binding / zinc ion binding	
At1g69120	AP1 (APETALA1); DNA binding / transcription factor	
At1g69180	CRC (CRABS CLAW); transcription factor	
At1g70510	KNAT2 (KNOTTED-LIKE FROM ARABIDOPSIS THALIANA 2); transcription factor	
At1g70830	MLP28 (MLP-LIKE PROTEIN 28)	
At1g75520	SRS5 (SHI-RELATED SEQUENCE 5)	
At1g76420	CUC3 (CUP SHAPED COTYLEDON3); transcription factor	
At1g77110	PIN6 (PIN-FORMED 6); auxin:hydrogen symporter/ transporter	
At2g02540	ATHB21/ZFHD4 (ZINC FINGER HOMEODOMAIN 4); DNA binding / transcription factor	
At2g23170	GH3.3; indole-3-acetic acid amido synthetase	
At2g26400	ARD/ATARD3 (ACIREDUCTONE DIOXYGENASE); acireductone dioxygenase [iron(II)-requiring]/ heteroglycan binding / metal ion binding	
At2g30370	CHAL (CHALLAH) Encodes a small, potentially secreted protein that acts as an inhibitor of stomatal production though likely not through direct interaction with the TMM receptor. It is homologous to known stomatal regulators EPF1 and EPF2	
At2g33880	WOX9 (STIMPY); transcription factor	
At2g42840	PDF1 (PROTODERMAL FACTOR 1)	
At2g45190	AFO (ABNORMAL FLORAL ORGANS); transcription factor	
At2g46870	NGA1 (NGATHA1); transcription factor	
At2g47460	ATMYB12/MYB12 (MYB DOMAIN PROTEIN 12); DNA binding / transcription activator/ transcription factor	
At3g04290	ATLTL1/LTL1 (LI-TOLERANT LIPASE 1); carboxylesterase	
At3g13960	AtGRF5 (GROWTH-REGULATING FACTOR 5)	
At3g19270	CYP707A4 (cytochrome P450	
At3g28500	60S acidic ribosomal protein P2 (RPP2C)	
At3g50870	MNP (MONOPOLE); transcription factor	
At3g57670	NTT (NO TRANSMITTING TRACT); nucleic acid binding / transcription factor/ zinc ion binding	
At4g00870	basic helix-loop-helix (bHLH) family protein	
At4g04890	PDF2 (PROTODERMAL FACTOR2); DNA binding / transcription factor	
At4g15440	HPL1 (HYDROPEROXIDE LYASE 1); heme binding / iron ion binding / monooxygenase	
At4g21750	ATML1 (MERISTEM LAYER 1); DNA binding / transcription factor	
At4g29030	glycine-rich protein	
At4g37750	ANT (AINTEGUMENTA); DNA binding / transcription factor	
At5g01370	unknown protein	
At5g02030	LSN (LARSON); DNA binding / transcription factor	
At5g03790	ATHB51/LMI1 (LATE MERISTEM IDENTITY1); DNA binding / sequence-specific DNA binding / transcription factor	
At5g11320	YUC4 (YUCCA4); monooxygenase	
At5g15310	AtMIXTA/AtMYB16 (myb domain protein 16); DNA binding / transcription factor	
At5g22500	acyl CoA reductase	
At5g23940	EMB3009 (EMBRYO DEFECTIVE 3009); transferase	
At5g28640	AN3 (ANGUSITFOLIA3)	
At5g53210	SPCH (SPEECHLESS); DNA binding / transcription factor	
At5g53950	CUC2 (CUP-SHAPED COTYLEDON 2); transcription factor	
At5g60910	AGL8 (AGAMOUS-LIKE 8); transcription factor	
At5g62165	AGL42 (AGAMOUS LIKE 42); transcription factor	

General abbreviations

αHA	antibody against HA
3´	three prime end of DNA fragment
5´	five prime end of DNA fragment
A	Adenine
AFAWE	Automatic functional annotaion in a distributed Web Services Environment
AP2/ERF	APETALA2/Ethylene-Responsive element binding Factor
apex/infl.	apex/inflorescence
Arabidopsis	*Arabidopsis thaliana*
At	*Arabidopsis thaliana*
At-TAX	*Arabidopsis thaliana* Tiling Array Express
AuxREs	auxin-response elements
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BLAST	Basic Local Alignment Search Tool
bp	base pairs
C	Cytosine
ChIP-chip	Chromatin immunoprecipitation followed by hybridisation to whole genome tilling arrays
ChIP-Seq	ChIP followed by sequencing
Col-0	Columbia
CZ	central zone
D	Aspartic acid
DNA	desoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
Drosophila	*Drosophila melanogaster*
E.coli	*Escherischia coli*
EDTA	Ethylendiamin-tetraacetat

EMBL-EBI	European Molecular Biology Laboratory- European Bioinformatics Institute
EMBOSS	European Molecular Biology Open Software Suite
F	Phenylalanine
FDR	false discovery rate
G	Glycine
G	Guanine
GA	Gibberelic Acid
GA	Gibberelic acid
Gabi	Genomanalyse im biologischen System Pflanze
GM	germination medium
GO	Gene Ontology
H3K27me3	tri-methylated lysine 27 at histone 3
H3K9me2	di-methylated lysine 9 at histone 3
HD-Zip III	class III homoedomain leucine zipper
IAA	indole-3-acetic acid
IDA	Inferred from Direct Assay
IPA	indole-3-pyruvic acid
ISS	Inferred from Sequence or Structural Similarity
KNOX	*K*NOTTED1-like homeobox
L	Leucine
L*er*	Landsberg *erecta*
LB	left border
LBD	Lateral organ Boundary Domain
LD	long day
LP	left border primer
M	Methionine
MADS	MCM1, AGAMOUS, DEFICIENS, and SRF
Mas 5.0	The Affymetrix®Microarray Analysis Suite 5.0
miR	micro RNA
mRNA	messenger RNA
N	Asparagine

NAC	NAM (NO APICAL MERISTEM)-ATAF1,2-CUC2
NASC	The Nottingham Arabisdopsis Stock Centre
NBT	nitro blue tetrazolium
ncRNA	non-coding RNA
OC	organising center
PBS	Phosphate buffered saline
PcG	Polycomb Group
PCR	polymerase chain reaction
PRC1-4	Polycomb Repressive Complexes
PREs	Polycomb group response elements
PZ	peripheral zone
qRT-PCR	quantitative real time polymerase chain reaction
R	Arginine
RMA	robust multi-array average
rms	root mean square
RNA	ribonucleic acid
RNAi	RNA interference
RP	right border primer
RZ	rib zone
SAIL	Syngenta Arabidopsis Insertion Library
SAM	shoot apical meristem
SD	short day
SEM	Scanning electron microscope
SET	SUVAR3-9/E(Z)/Trithorax
SIGnAL	Salk Institute Genomic Analysis Laboratory
SSC	saline sodium-citrate buffer
T	Thymine
TAIR	The Arabidopsis Information Resources
TE	Tris/EDTA
UIM	ubiquitin interaction motif
V	Valine

Gene names

ABC	ATP-binding cassette
ABI3	ABSCISIC ACID-INSENSITIVE 3
AG	AGAMOUS
ANT	AINTEGUMENTA
AP1-3	APETALA1-3
ARFs	AUXIN RESPONSE FACTORS
AS1	ASYMMETRIC LEAVES 1
AS2	ASYMMETRIC LEAVES 2
AUX/IAA	Auxin/Indole-3-acetic acid inducible
AUX1	AUXIN1
BB	BIG BROTHER
BP	BREVIPEDICELLUS
BPE	BIGPETAL
CAL	CAULIFLOWER
CDI	cyclin-dependent kinase inhibitor
CDK	cyclin-dependent kinase
CER	ECERIFERUM1
CESAs	cellulose synthase catalytic subunits
CLF	CURLY LEAF
CLV1-3	CLAVATA
CLV2	CLAVATA2
CO	CONSTANS
COP1	CONSTITUTIVE PHOTOMORPHOGENESIS 1
CRY1/2	CRYPTOCHROME 1/2
CSD	copper/ zinc (Cu/ Zn) superoxide dismutase
CSLA10	CELLULOSE SYNTHASE LIKE A10
CUC2	CUP-SHAPED COTYLEDON 2
DA1	DA means“larg” in Chinese
DOT5	DEFECTIVELY ORGANIZED TRIBUTARIES 5

DPAs	Development related PcG Targets in the Apex
DVL21	DEVIL 21
E(Z)	Enhancer of Zeste
E-GFP	endoplasmic reticulum targeted GFP
EMF1	EMBRYONIC FLOWER1
EMF2	EMBRYONIC FLOWER 2
ESC	Extra Sex Combs
FIE	FERTILIZATION INDEPENDENT ENDOSPERM
FIS2	FERTILIZATION INDEPENDENT SEED 2
FKF1	FLAVIN-BINDING, KELCH REPEAT, F-BOX PROTEIN 1
FLC	FLOWERING LOCUS C
FMO	FLAVIN-MONOOXYGENASE
FRI	FRIGIDA
FT	FLOWERING LOCUS T
FUS3	FUSCA3
GFP	Green Fluorescent Protein
GI	GIGANTEA
GS-OX	GLUCOSINOLATE S-OXYGENASE
H1	histone 1
H2A	histone 2A
H2B	histone 2B
H3	histone 3
H4	histone 4
HA	hemagglutinin
HSI	High-level expression of sugar-inducible gene
JAG	JAGGED
KNAT2	KNOTTED-LIKE FROM ARABIDOPSIS THALIANA 2
KRP1/2	Kip-related proteins
LAX1-3	LIKE AUX1
LEC2	LEAFY COTELYDON 2
LFY	LEAFY
LOF1	LATERAL ORGAN FUSION1

MEA	MEDEA
MSI	Multicopy Suppressor of IRA
MSI1-5	MSI five homologs
NAM	NO APICAL MERISTEM
NGA1-4	NGATHA 1-4
PC	Polycomb
PDF2	PROTODERMAL FACTOR2
PH	Polyhomeotic
PHB	PHABULOSA
PHV	PHAVOLUTA
PhyA/B	PHYTOCHROME A/B
PID	PINOID
PIN1	PIN-FORMED 1
PP2A	protein phophatase 2A
PSC	Posterior Sex Combs
RAV	Related to ABI3/VP1
REM	Reproductive Meristem
RING	RING FINGER PROTEIN
RTF	ROTUNDIFOLIA
SCE	Sex Combs Extra
SE	SERRATE
SHI	SHORT INTERNODES
SMZ	SCHLAFMÜTZE
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
SPA1	SUPRESSOR OF PHYA-105-1
SPL	SQUAMOSA PROMOTER BINDING LIKE
SRF	serum response factor
SRS8	SHORT INTERNODES RELATED SEQUENCE 8
STM	SHOOTMERISTEMLESS
STY1/2	STYLISH 1/2
Su(Z)12	Suppressor of Zeste 12
SVP	SHORT VEGETATIVE PHASE

SWN	SWINGER
TAA1	TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS 1
TAR2	TRYPTOPHAN AMINOTRANSFERASE RELATED 2
TCP	TEOSINTE BRANCHED1/CYCLOIDEA/PCF
TEM1	TEMPRANILLO 1
TEM2	TEMPRANILLO 2
TFL2/ LHP1	TERMINAL FLOWER 2/LIKE HETEROCHROMATIN PROTEIN 1
UFO	UNUSUAL FLORAL ORGANS
VP1	VIVIPAROUS1
VRN1	VERNALIZATION 1
VRN2	VERNALIZATION 2
WOX9	WUSCHEL-RELATED HOMEOBOX 9
WUS	WUSCHEL
YUC	YUCCA

Acknowledments

I would like to thank ...

Franziska Turck for the opportunity to work on this interesting project in her lab and for the helpful discussion and supervision.

Prof. George Coupland for the opportunity to do my PhD in his department and for the helpful discussion and comments to my project.

Prof. Wolfgang Werr for being my second examiner and for supporting the project with expertise on embryo *in situ* hybridisation and by the provision of *CUC* mutants.

Prof. Martin Hülskamp for being the head of my thesis committee.

Peter Huijser for being my second supervisor and supporting my work with many helpful comments and ideas.

Everybody in the Turck Group for the nice time, support and explanations. Especially **Julia Reimer** for her helpful advice and the Perl-lessons, **Sara Farrona** for a lot of useful comments and the help to detect abnormal phenotypes, **Jessika Adrian** for answering so many questions during my first months in the lab and **Valentina Strizhova** for help with DNA preparations and crosses.

Ulrike Göbel and **Anika Jöcker** for help with bioinformatic questions.

Elmon Schmelzer and **Rainer Franzen** for help with the electron microscopy.

Bruno Hüttel for performing the microarray hybridisation and for helpful advice for experiment design and data analysis.

Stefano Torti, **Maida Romera Branchat**, **Coral Vincent** and **Melanie Cole** for help and advice on *in situ* experiments.

Christopher Kraus and **Lisa Stephan** for help in the lab.

The Gardeners for taking care of my plants.

Everybody in the Coupland Department for the nice atmosphere.

Markus Berns for the fun in the old office, all the nice discussions and for sharing the ups and downs of the PhD life.

Christian Kröner for his patience, computer help and for always being there for me.

My Parents and **my Grandma** for their love and for all the support.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. George Coupland betreut worden.

Köln, den 6. Dezember 2010 _____

Julia Carina Engelhorn

Teilpublikationen

Engelhorn, J. and Turck, F. (2010). Metaanalysis of ChIP-chip data. *Methods Mol Biol*, **631**, 185-207.

Lebenslauf

Persönliche Angaben

Name	Julia Carina Engelhorn
Geburtsdatum	17. August 1982
Geburtsort	Haan
Nationalität	deutsch
Familienstand	ledig

Ausbildung

2001	Abitur, Röntgen-Gymnasium, Remscheid, Deutschland
2001–2005	Studium der Biochemie, Universität Bayreuth, Deutschland Studienschwerpunkte: Biochemie, Genetik, Mikrobiologie, Biophysikalische Chemie und Pflanzenphysiologie
2005	Diplomarbeit am Lehrstuhl für Struktur und Chemie der Biopolymere, betreut von Prof. Dr. P. Rösch, "Expression und Reinigung von rekombinanter Norcoclaurinsynthase (NCS)"
2005	Abschluss: Diplom Biochemie
Oktober 2007	Beginn der Doktorarbeit am Max-Planck-Institut für Züchtungsforschung, Köln in der Arbeitsgruppe von Dr. Franziska Turck.

Anstellungen

2006–2007	Clinical research associate, Winicker Norimed GmbH, Nürnberg